

MODULE 1- LECTURE 1

ROLE OF GENES WITHIN CELLS, GENETIC CODE, GENETIC ELEMENTS THAT CONTROL GENE EXPRESSION

1-1.1 Introduction:

The main distinction between a living and non-living entity is the ability to replicate and reproduce similar offsprings. Nucleic acid molecules (DNA and RNA) present in a living organism acts as a genetic template to pass the hereditary information from one generation to the next. Nucleic acid molecules are organised as genes which code for a particular phenotype via specific proteins and the expression of a gene is regulated by both external and internal factors which aid the developmental process of an organism. This relation between genes and proteins forms the “central dogma of life”.

1-1.2 Gene:

A gene can be defined as the region of DNA (or RNA in case of virus) that controls a discrete hereditary characteristic, usually corresponding to a single protein or RNA. This includes the entire functional unit, encompassing coding (exons) and noncoding sequences (introns and regulatory sequences).

- Exons and introns which represent the coding and noncoding regions are present in a eukaryotic gene. Introns are absent in prokaryotes.
- The introns are removed by splicing and the exons are translated in tandem to yield the functional polypeptide that further undergoes post translational modification to become functional. These functional polypeptides (proteins) are targeted to various organelles in the cell or exported out of the cell for carrying out various intracellular and extracellular processes respectively.

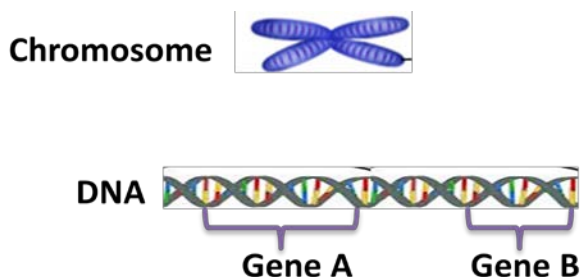


Fig 1-1.2: Organization of genes in DNA of a chromosome

1-1.3 Genome:

Genome is the complete set of genetic information of a cell or an organism; in particular, the complete sequence of DNA/RNA that carries this information. In diploid organisms, it refers to the haploid set of chromosomes present in a cell. Depending on its localization, genome may be nuclear or organellar. Organellar genomes are again of two types: mitochondrial and chloroplast genome. Genome size of organisms differs significantly between different species. The size of the genome governs the size and complexity of an organism. However, many small sized organisms, in fact have bigger genomes than their larger counterparts.

Various organisms have different sized genome as can be seen in the table below.

Species	Organism	Genome Size (Mb)
<i>Triticum aestivum</i>	Plant	16000
<i>Homo sapiens</i>	Mammal	3200
<i>Arabidopsis thaliana</i>	Plant	125
<i>Drosophila melanogaster</i>	Insect	180
<i>Caenorhabditis elegans</i>	Nematode worm	97
<i>Saccharomyces cerevisiae</i>	Yeast	12.1
<i>Escherichia coli</i>	Bacterium	4.64
<i>Haemophilus influenzae</i>	Bacterium	1.83
<i>Mycoplasma genitalium</i>	Bacterium	0.58

- The genome contains all the genes present in the nucleus of a cell. Gene varies in size from a few hundred DNA/RNA bases to more than few thousand bases.
- The haploid set of chromosome contains the total genome of the organism.
- The bacterium *Mycoplasma genitalium* has a small genome size of 0.58Mb and the plant *Triticum aestivum* has a large genome size of 16000Mb. The genome size in human is 3200Mb.

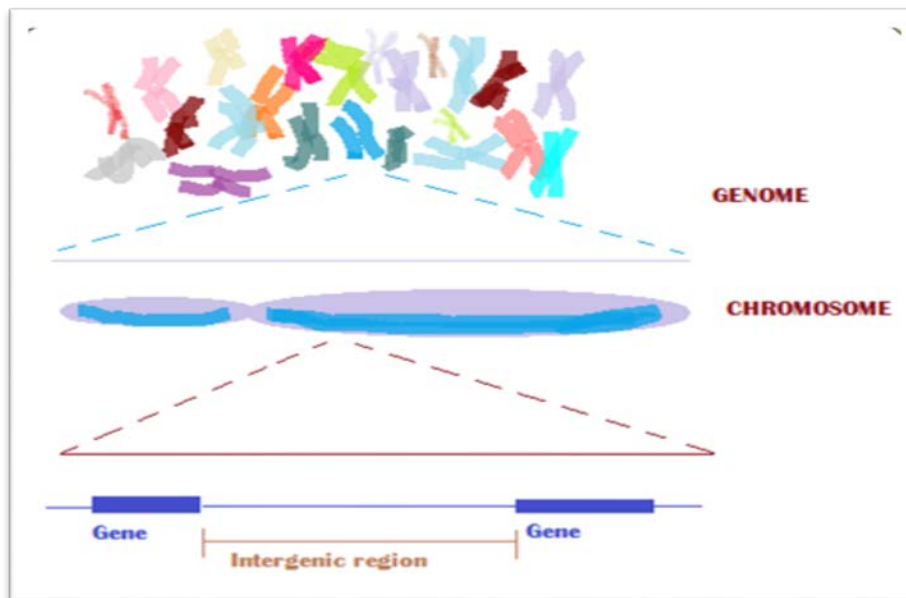


Fig 1-1.3: Illustration of genome, from genome to a chromosome, from a chromosome to gene

1-1.3.1 Role of genes within cells:

- Genes contain the instructions for each cell to make proteins and RNAs. Genes are made up of DNA fragments.
- Within the cell the DNA performs two tasks:
 - Act as information repository including instructions in making the component molecules of the cells.
 - Pass on the information to the next generation.
- The mere presence of DNA does not implicate a cell to be alive and functional. Mammalian red blood cells (RBCs) discard nucleus during developmental process and thus lacks DNA in mature state.

- Genes are transcribed to RNA which are processed to various forms like mRNA, tRNA, rRNA etc. mRNA are translated to proteins depending on the regulatory signals. tRNA and rRNA serve as the components of translational machinery.
- New functions of RNA are also being discovered like regulatory (miRNA, siRNA etc) and catalytic (ribozymes) functions.
- Proteins are structural components of cells, enzymes, hormones, various signalling molecules, receptors and other factors which are involved in performing the chemistry of life and are essential for the normal body function, for example, sugar conversion to energy and metabolite (small molecules) production in cell.

1-1.4 Genetic code:

The genetic code is the set of instructions that translates the information encoded in genetic material (mRNA or DNA sequences) into proteins (amino acid sequences) by living cells.

- The genetic code is a triplet code (i.e. a group of three adjacent nucleotides) called codon. This three nucleotide codon in nucleic acid sequence specifies a single amino acid. However, genetic code in human mitochondria differs from the standard nuclear genetic code.
- Epigenetic effects, however is not stored using the genetic code. Besides all organisms have DNA that contains regulatory sequences, intergenic segments, chromosomal structural areas and other non-coding DNA that can contribute greatly to phenotype.
- These codons are always written with the 5'-terminal nucleotide to the left.
- The code is unambiguous i.e. each triplet specifies only a single amino acid.
- The genetic code is degenerate i.e. more than one triplet codon can code for a single amino acid. (61 codons code for 20 amino acids)

- Three codons do not specify any amino acid but acts as terminal sites (stop codons), signalling the end of protein coding sequence. They are namely UAA (Ochre), UAG (Amber) and UGA (Opal).
- AUG is an initiation codon that signals the initiation of translation and also codes for methionine. In some mRNA, GUG and UUG also act as initiation codon.
- The genetic codes for each 20 amino acids were defined by pioneering works of Marshall W. Nirenberg and Har Gobind Khorana in 1964.

		Second letter→					
First letter→		U	C	A	G		Third letter→
	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA } Stop UGG } Trp	U C A G	
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G	
	A	AUU } Ile AUC } AUA } Met AUG }	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G	
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G	

Table 1-1.4: Genetic code

1-1.5 Gene expression:

Let us now understand the central dogma of life. You now know that most of the genes contain the information needed to make functional molecules called proteins. However, some genes produce other molecule (tRNA, rRNA, microRNA etc) that assist or regulates the protein expression and assembly. These complex events within each cell consist of two main steps: transcription and translation. In prokaryotes where there is no nuclear membrane, both transcription and translation occur in the cytoplasm whereas in eukaryotes, transcription occurs inside the nucleus and the translation occurs in the cytoplasm.

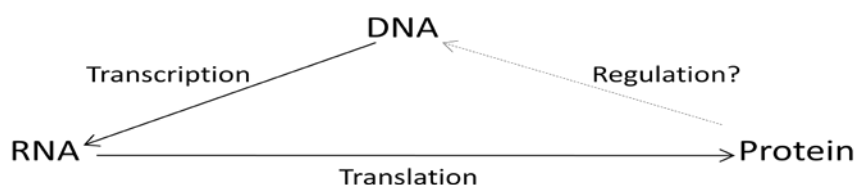


Fig1-1.5: Central dogma of life

1-1.5.1 Events involved in gene expression:

The events involved in genome expression in higher organisms with respect to protein coding genes are discussed below. The genes that give rise to non-coding RNAs are transcribed and processed but are never translated.

- **Access to gene:** Genes are inaccessible as they are buried deep within the highly packaged chromosomes. The initial step involves a preparative process that opens the chromatin structure and positions of the nucleosome in the region of genome containing active genes.
- **Formation of transcription initiation complex** involves the assembly of a set of proteins into a complex that copy DNA into RNA. This is a highly regulated process as the transcription initiation complex must be constructed at the precise position in the genome, adjacent to active genes to form a RNA copy.

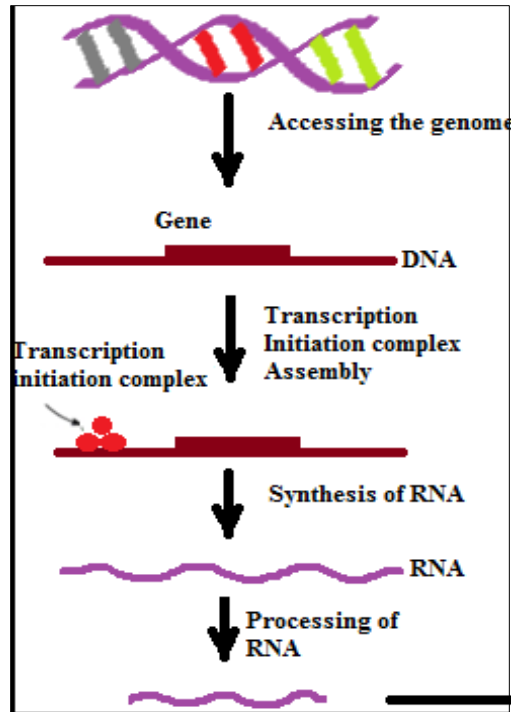


Fig: 1-1.5.1: Steps involved in Gene Expression

- **RNA Synthesis** involves the transcription of a gene into RNA molecule and it occurs in the nucleus.
- **RNA Processing** comprises of post transcriptional modification/alterations of the RNA molecule and its chemical structure required for the RNA to be translated into protein or non-coding RNA (rRNA, tRNA, miRNA). RNA splicing (deletion of introns and combination of exons), 5' capping, polyadenylation etc are commonly occurred RNA processing steps in eukaryotes. However, prokaryotic organisms do not have a well developed RNA processing machinery.
- **Degradation of RNA** is the controlled turnover of RNA molecules and should not be viewed simply as a mean of getting rid of unwanted RNAs. It determines the makeup of the transcriptome and is considered as an important step in genome expression. Different ribonucleases (RNases) play the prime role in this process and multiple cofactors like small RNA (siRNA, miRNA etc), molecular chaperons (Lsm1-7, Lsm2-8, Hfq etc) regulate this process.

- **Protein synthesis** is initiated after the assembly of the translation initiation complex near the 5' termini of a mRNA molecule. It involves translation of RNA molecules into proteins.
- **Protein folding and protein processing** may occur together after protein synthesis. Post translation events like folding involve the protein attaining its correct three dimensional configuration. Processing (phosphorylation, glycosylation, carboxylation etc) involves the modification of the protein by addition of chemical groups and removal of one or more functional units of the protein.

1-1.5.2 Types of gene expression:

- **Constitutive expression:** Housekeeping genes are essential and necessary for sustaining life, and are therefore continuously expressed. *gapdh* (glyceraldehydes 3 phosphate dehydrogenase), *sdha* (succinate dehydrogenase) etc are human housekeeping genes which are expressed throughout the development.
- **Induction and repression:** The expression levels of some genes fluctuate in response to external signals. Also, under a certain situation, some genes show higher expression level, while others show lower expression levels. The former is called induced expression and the latter is called repressed expression.

1-1.6 Control of gene expression:

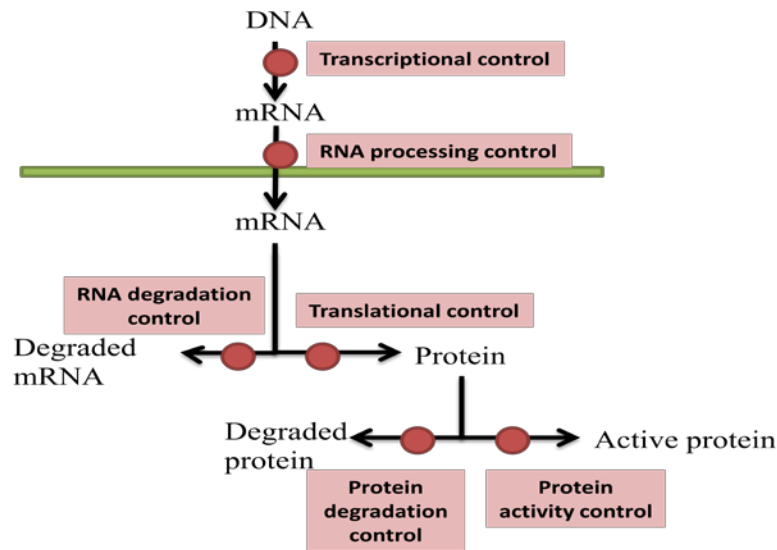


Fig1-1.6: Points of control of gene expression

The eukaryotic gene expressions in a cell are controlled at about six stages as shown in the figure 1-1.6.

- Transcriptional control decides when and how frequently a given gene be transcribed. It enables the cell to check any discrepancy in the quantity of transcriptomes generated.
- RNA processing regulates how the RNA transcript is spliced or otherwise processed (applicable only to eukaryotes), which have split genes consisting of exons and introns.
- Translational control decides which mRNAs in the cytoplasm are to be translated by ribosomes.
- mRNA degradation control selectively destabilize some mRNA molecules in the cytoplasm.
- Protein activity control selectively activates, inactivates, degrades, or compartmentalizes specific protein molecules after they have been synthesized. (applicable only to eukaryotes)

1-1.6.1 Regulatory elements:

Gene expression is a complex multi-step process. Amongst various steps involved, transcription initiation is a vital key point in controlling gene expression.

Fundamental elements that regulate the process of transcription are:

- cis-acting elements comprising of special DNA sequences.
- trans-acting elements comprising of regulatory proteins.
- DNA –protein interaction.
- Protein-protein interaction
- RNA polymerase

► **cis-acting elements:**

In Latin “cis”, means "same side as". Cis acting elements are thus a region of DNA or RNA that regulates the expression of genes located on the same molecule. The cis-regulatory elements are often binding sites for one or more trans-acting factors. Cis-element may be located upstream to the coding sequence of the gene it controls (in the promoter region or further upstream 5'), in an intron, or 3' to the gene's coding sequence, either in the untranslated or untranscribed region.

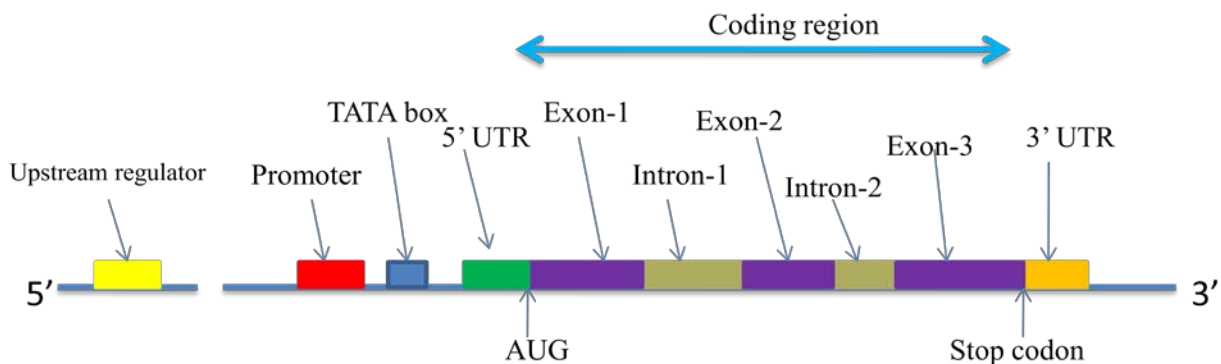


Fig1-1.6.1: Structure of eukaryotic gene with upstream regulatory elements

Examples of cis-acting elements:

Prokaryotic systems:

- Promoter is a DNA sequence where RNA polymerase binds to initiate transcription. There are two promoter sequences in prokaryotic systems known as -10 (Pribnow box or Pribnow-Schaller box) and -35 sequences.
- Operator is regulatory sequence of DNA located immediately upstream of the structural gene that controls transcription of an operon.
- Inducers are located upstream of promoter region.
- Downstream regulatory sequences comprise of GC-rich inverted repeats followed by four adenine (A) residues signal the termination of transcription.

▶ **Trans-acting elements:**

The protein factors which regulate the expression of gene by binding to cis acting DNA sequence are termed as trans-acting elements. Trans-acting molecules generally have two domains: DNA binding domain (which binds to cis elements) and protein binding domain (required for activation or suppression of transcription). Transcription initiation is a tightly regulated process controlled by trans-acting elements both in prokaryotes as well as eukaryotes.

▶ **DNA – Protein interactions:**

Gene regulatory proteins and the transcription factors are capable of binding to the DNA based on the interaction of amino acids of the protein with the nucleotides of the DNA. The regulation is implemented through various interactions between cis-acting elements and trans-acting factors. Examples of some DNA binding proteins are:

- Eukaryotic TATA-binding protein
- σ subunit of bacterial RNA polymerase etc.

There are four types of structures of DNA binding proteins,

1. Zinc finger proteins
2. Helix loop Helix protein
3. Leucine zipper proteins
4. Homeodomain proteins

► **Protein – Protein interactions:**

Protein-protein interaction is present in both prokaryotes and eukaryotes. The external signals affect the gene expression with the help of such interaction. Proteins interact with each other to form a homo or hetero-dimers before binding to DNA molecules. Eg. transcription initiation factors interacts with TATA box binding proteins (TBP) to activate transcription of a gene.

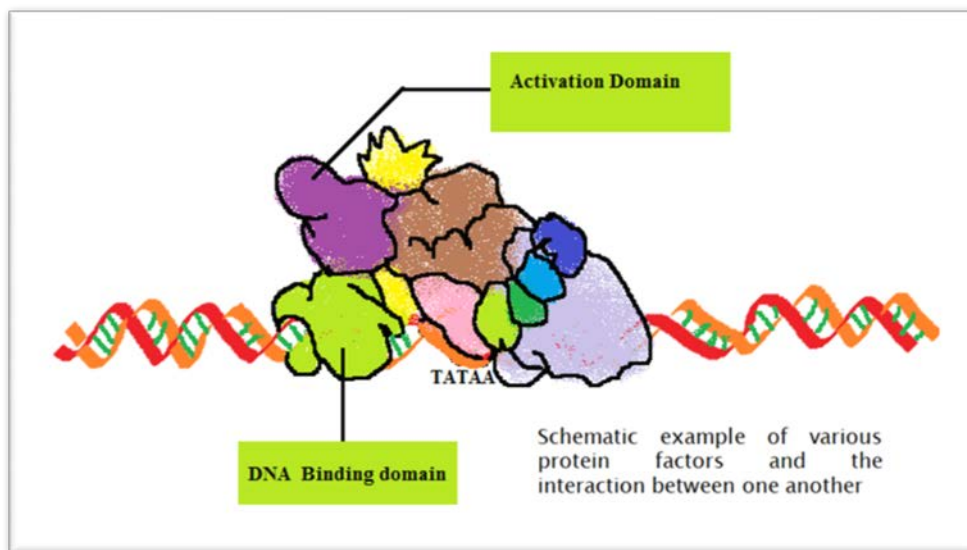


Fig1-1.6.2: Interaction between various protein factors

► **RNA Polymerase:**

- A single RNA polymerase is responsible for transcribing all types of RNA in prokaryotic system.
- However, eukaryotes have three different RNA polymerases, which have been found to specialize in the synthesis of various types of RNA:
 - RNA polymerase I (Pol I) - transcribes rRNA (ribosomal RNA) genes.

- RNA polymerase II (Pol II) - transcribes protein-coding genes or mRNA (messenger RNA).
- RNA polymerase III (Pol III) - transcribes other functional RNA genes (e.g., tRNA).
- In eukaryotes, transcription occurs inside the nucleus. All the enzymes responsible for translation are present in the cytosol therefore the transcripts formed then move out of the nucleus through nuclear pores into the cytosol (the liquid phase of the cytoplasm), where translation occurs. Organelle genomes (like the mitochondrial genome) are transcribed within the organelle (the mitochondria) and translation is also within the organelle (the mitochondria).
- Since prokaryotes have no nucleus, the step involving the movement of transcripts from nucleus to cytoplasm does not take place, and translation can take place immediately in the cytoplasm, directly on the growing transcript.

Bibliography:

- Alberts B, Johnson A, Lewis J, Raff M., Roberts K., Walter P. 2002. *Molecular Biology of the Cell* (4th edition); New York: Garland Science.
- Brown T.A. 2002. *Genomes* (2nd edition); Oxford: Wiley-Liss.
- Cooper G.M. 2000. *The Cell: A Molecular Approach* (2nd edition); Sunderland (MA): Sinauer Associates.
- Dudek R.W. 2010. *High-Yield Cell and Molecular Biology* (3rd edition); Lippincott Williams and Wilkins.
- Griffiths A.J.F, Miller J.H., Suzuki D.T, Lewontin R.C., Gelbert W.M. 2000. *An Introduction to Genetic Analysis* (7th edition); New York: W. H. Freeman.

MODULE 1-LECTURE 2

METHOD OF CREATING RECOMBINANT DNA MOLECULES

1-2.1 Recombinant DNA:

- Single chimeric DNA formed by combining two or more different fragments of DNA from diverse organisms is generally called as recombinant DNA and the method applied to create recombinant DNA is called recombinant DNA technology.
- The organism, from which the candidate DNA is isolated, is called **Donor organism**. The organism which will accept the foreign gene is called **Host organism**.
- Genetic material from one organism is selected and then artificially introduced to a host organism. If the foreign recombinant DNA integrates into the host genome, it gets replicated along with the genome and then express the foreign protein.
- Paul Berg, Herbert W. Boyer and Stanley N. Cohen are the pioneers of recombinant DNA technology (early 1970).
- A hybrid of the *SV40* mammalian DNA virus genome and phage λ was one of the recombinant DNA molecules to be first engineered.
- There are three approaches to make recombinant DNA:
 1. Transformation
 2. Non- bacterial transformation/transfection
 3. Phage introduction/transduction

Transformation:

Transformation is direct uptake of exogenous DNA via cell membrane leading to incorporation into the host DNA. It is commonly occurred in bacteria. Transformation requires different tools of molecular biology to insert foreign DNA into the host. For example, **vector** to carry the foreign DNA to the host; **restriction enzymes** to cut the DNA in specific site; **ligase** to join two DNA molecule etc.

Non-bacterial transformation/transfection:

The process of foreign DNA uptake by host cell driven by mechanical or chemical factors is classified under non-bacterial transformation, also termed as transfection. Different methods of non-bacterial transformation are microinjection, liposome mediated transformation, biolistics etc.

Phage introduction/transduction:

Phage vector is used to carry and replicate foreign DNA inside the bacterial host system. The phage DNA inserts into the host chromosome by recombination. Phage λ had short regions of single-stranded DNA with complementary base sequences called “cohesive” (*cos*) sites. Base pairing between the complementary *cos* sites allows the linear genome to form a circle within the host bacterium. Circularized viral genome can be integrated into the bacterial genome by homologous recombination between *attP* site of viral genome and *attB* site of bacterial genome.

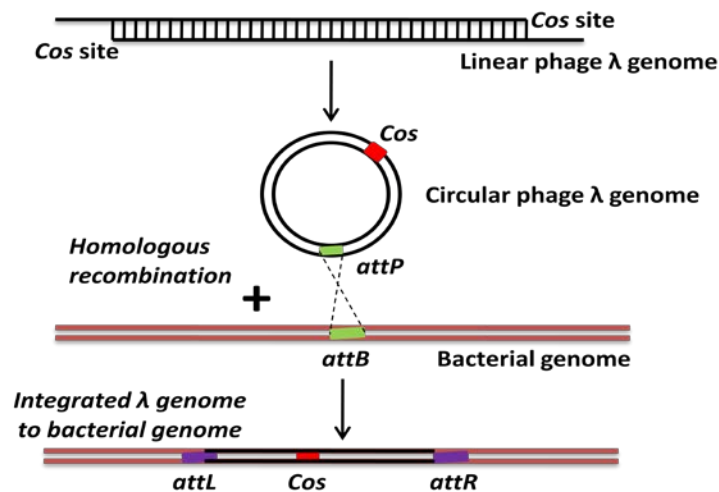


Fig 1-2.1: Bacteriophage λ genome circularization and recombination with bacterial genome

1-2.2 Methods involved in Recombinant DNA Technology:

Molecular cloning is a process for creating recombinant DNA and generally involves the following steps:

- (1) Selection of a cloning vector
- (2) Selection of a host organism
- (3) Preparation of a vector DNA
- (4) Preparation of DNA to be cloned
- (5) Creation of recombinant DNA vector (having foreign DNA)
- (6) Introduction of recombinant vector into host organism
- (7) Selection of clones having insert vector.
- (8) Screening and multiplication of recombinant clones with desired DNA inserts.

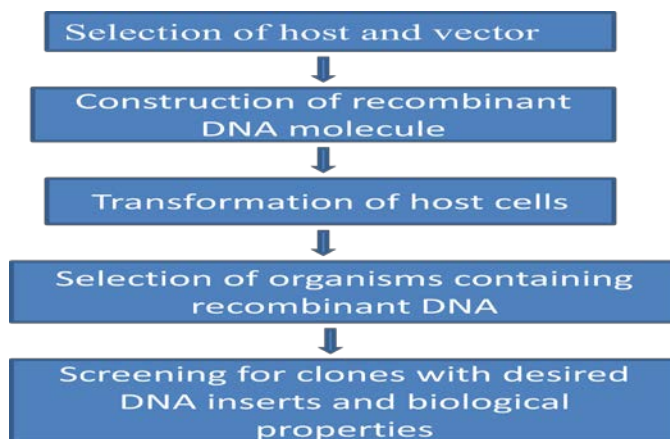


Fig 1-2.2: Steps in Gene Cloning

1-2.2.1 Choice of host organism:

A good host should have the following properties:

- Easy to grow and transform.
- Do not hinder replication of recombinant vector.
- Do not have restriction and methylase activities.
- Deficient in recombination function so that the introduced recombinant vector is not altered.
- Easily retrievable from the transformed host.

Various hosts are used in rDNA technology depending on the goal: For example bacteria, yeast, plant cells, animal cells, whole plants and animals.

- **Prokaryotic systems** such as *E. coli* are commonly used due to various advantages like, They have-
 - Well studied expression system,
 - Compact genome,
 - Versatile,
 - Easy to transform,
 - Widely available, and
 - Rapid growth of recombinant organisms with minimal equipment.

Only disadvantage is that they lack post-translational modification (PTMs) machinery required for eukaryotic proteins.

- **Eukaryotic systems** are difficult to handle in contrast to bacterial hosts. They are favoured for expression of recombinant proteins which require post translational modification and only if they can grow easily in continuous culture.

1-2.2.2 Choice of vector:

Vector is an autonomously replicating (inside a host cell) DNA molecule designed from a plasmid or phage DNA to carry a foreign DNA inside the host cell. Transformation vectors are of two types:

- Cloning vector is used increasing the number of copies of a cloned DNA fragment.
- Expression vector is used for expression of foreign gene into a protein.
- If a vector is designed to perform equally in two different hosts, it is called a shuttle vector.

Properties of an ideal vector: A good vector should have the following characteristics:

- Autonomously replicating i.e. should have ori (origin of replication) region.
- Contain at least one selectable marker *e. g.* gene for antibiotic resistance.

- May contain a scorable marker (β -galactosidase, green fluorescent protein etc.)
- Presence of unique restriction enzyme site.
- Have multiple cloning sites.
- Preferably small in size and easy to handle.
- Relaxed control of replication to obtain multiple copies.
- Presence of appropriate regulatory elements for expression of foreign gene.
- High copy number

The selection of a suitable vector system depends mainly on the size limit of insert DNA and the type of host intended for cloning or expression of foreign DNA.

List of different vectors

- **Plasmids** are circular DNA molecules that exist independently of chromosomal DNA and can replicate autonomously. Plasmids carry one or more genes which mostly code for useful characteristic of host. All plasmids have sequence that can act as origin of replication. Plasmids of different sizes and possessing different copy number are present.
- **Phage vectors** are consist of mainly DNA molecule (sometimes RNA); that carries large number of genes and are surrounded by a protein coat called as capsid. They can be used as vehicles for carrying DNA insert after modification to remove pathogenic genes and minimizing the size.
- **Cosmids** are hybrid between a phage DNA and bacterial plasmid. They have *cos* sites which are essentially required for packaging lambda (λ) into phage protein coat. They can carry large DNA insert.
- **Fosmids** are cosmid like plasmid but they are based on F-plasmid.
- **Phagemids** are plasmids having a part of M13 genome.
- **Artificial chromosomes** are artificially constructed DNA construct used for transferring DNA.

1-2.2.3 Preparation of vector DNA:

The vector DNA is cleaved by restriction endonucleases at the site where foreign DNA is desired to be inserted. The restriction enzyme is selected to generate a configuration at the cleavage site compatible with the ends of the foreign DNA. This can be achieved either by cleaving the foreign DNA and vector DNA with the same restriction enzyme or by adding adaptors/ linkers to both the ends of the insert DNA.

1-2.2.4 Preparation of DNA to be cloned:

DNA to be cloned can be obtained by:

1. Cutting using restriction enzyme from genomic or organellar DNA,
2. PCR based amplification,
3. Chemical synthesis.

The DNA to be cloned is isolated and treated with restriction enzymes to generate random fragments with ends capable of being linked to those of the vector. While choosing the restriction enzyme to cut the desired gene, care should be taken so that the restriction enzyme does not cut in the middle of the gene, but only at the ends. PCR based methods are used to obtain DNA segments, using either genomic DNA or mRNA as template sequences through reverse transcription. Short length sequences can be artificially synthesized *in vitro*. If necessary, linkers or adapters containing desired restriction sites are added to create the ends which are compatible with the vector. The complementary sticky ends result in an efficient ligation due to the formation a stable structure.

1-2.2.5 Creation of recombinant DNA by ligation

- The vector DNA, foreign DNA and DNA ligase enzyme are added together at appropriate concentrations which results in the covalent linkage between the ends of DNA fragments.
- DNA ligase recognizes the ends of linear DNA molecules and gives a complex mixture of DNA molecules with randomly joined ends.
- The resulting recombinant DNA vector is then introduced into the host organism.

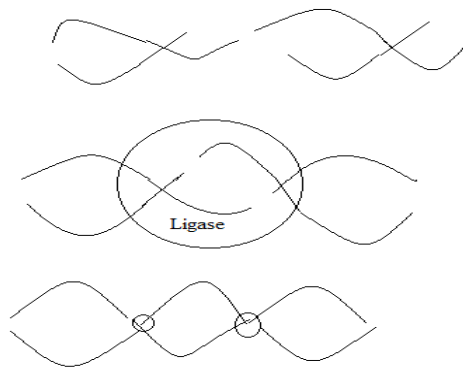


Fig 1-2.2.5: Ligation using ligase enzyme

- In addition to desired recombinant DNA, complex mixture containing self ligated vector DNA, foreign DNA linked with other sequences and several other combinations of vector and foreign DNA also appear in the reaction mixture.
- Sorting of the complex mixture is done by agarose gel electrophoresis based on size of the recombinant vector.

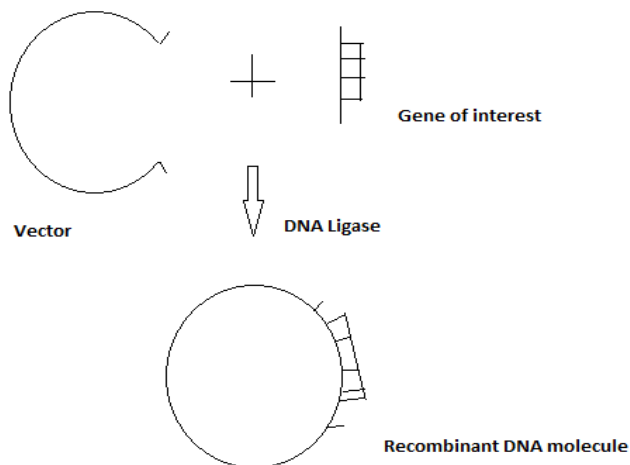


Fig 1-2.2.5.1: Preparation of recombinant DNA

1-2.2.6 Introduction of recombinant DNA into host organism:

- For the propagation of a cloned gene, the recombinant DNA molecules have to be introduced into a host.
- Numerous methods of gene transfer are available to meet the diverse requirement and compatibility with the host (e.g. transformation, transduction, transfection, electroporation etc.).
- **Transformation** is the process in which microorganisms are able to take up the DNA from their surrounding via plasma membrane and express it. Cells should be competent to take up the foreign DNA.
- DNA transfer into mammalian cells cultured in vitro using non-viral vectors is termed as **transfection**.
- **Transduction** is the process of transfer of DNA molecule using viruses.
- Both transformation and transfection requires preparation of the cells through a specific growth condition and chemical treatment process that varies with the specific species and cell types to be used. For example – Calcium chloride is used for preparation of competent *E.coli* cells.

- Electroporation uses electrical pulses to create transient holes in the cell membrane through which DNA is translocated across the cell membrane. Cell wall has to be previously removed in plants to increase the rate of transfer. Electroporation is usually done by two methods:
 - High voltage for short time,
 - Low voltage for long time.
- Electroporation and transduction are very efficient methods to transfer DNA into cells.

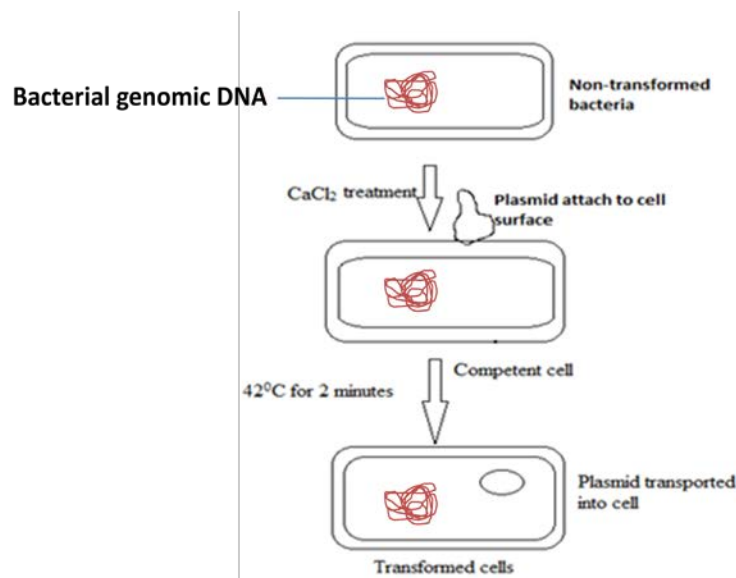


Fig 1-2.2.6: Preparation of competent cells

1-2.2.7 Selection of host cells/ organism containing vector sequences:

Selection of the transformed cells from the non-transformed population is done by using selectable marker genes that confers resistance to antibiotics. Hence, cells only having the vector with the resistance gene for the antibiotic would grow in the selection media containing the antibiotic (ampicillin, tetracycline etc.); while the non-transformed cells would die.

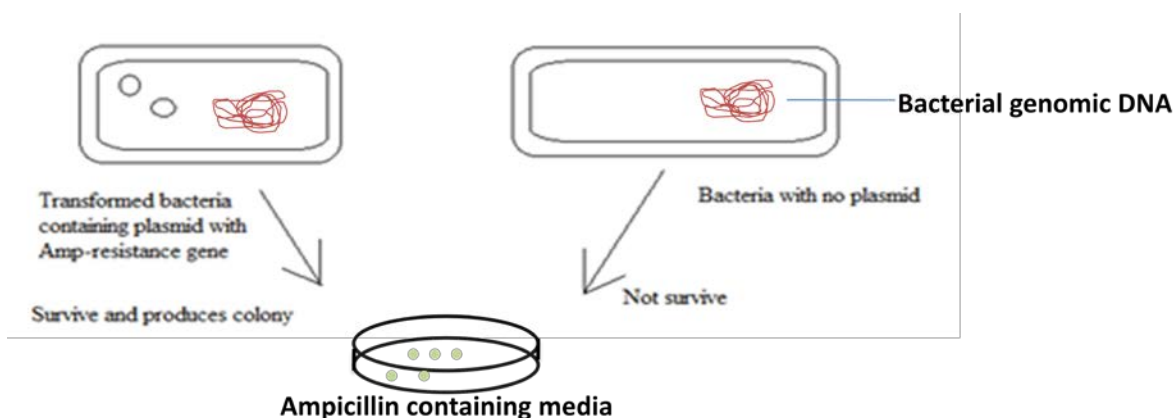


Fig 1-2.2.7: Selection of transformed cells

1-2.2.8 Screening clones having desired DNA inserts with the help of biological properties:

- After selection of colonies having the vector, the next step is to screen the colonies having the recombinant vector (vector containing foreign DNA insert).
- Bacterial cloning vectors (e.g. pUC19, pGEM vectors) use the blue-white screening system based on *lacZ* system to distinguish transgenic cells from those that contain the parental vector (i.e. vector DNA with no recombinant sequence inserted). The recombinant colonies are grown in presence of X-gal.
- In these vectors, foreign DNA is inserted into a sequence that encodes an essential part of beta-galactosidase (an enzyme which cleaves galactose). Its activity results in formation of a blue-colour colony on the culture medium.

- Insertion of the foreign DNA into the beta-galactosidase coding sequence disrupts the function of the enzyme, and colonies containing recombinant plasmids give no blue colour (white).
- Using this colour phenotype, transgenic bacterial clones can be easily identified from those that do not contain recombinant DNA.

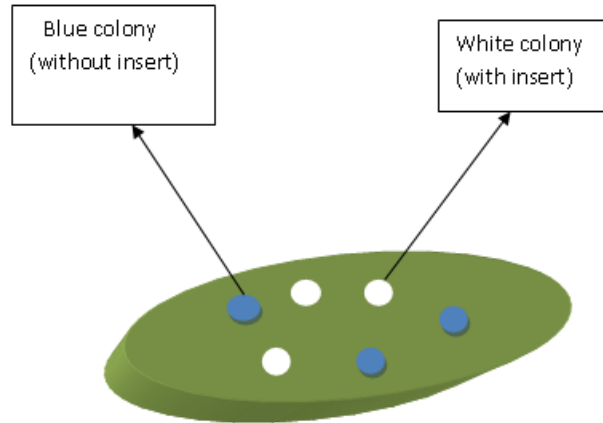


Fig 1-2.2.8: Screening for clones with desired DNA inserts

Insertional inactivation of antibiotic gene can also be used for the selection of recombinant cells.

- A vector is chosen where restriction sites are available for cloning within the antibiotic gene. Insertion of a foreign gene in the restriction site will lead to the loss of activity of the selectable marker (antibiotic) gene. For example-pBR322 have several restriction sites. *Bam*HI cuts at a one position within genes that code for tetracycline resistance. Thus recombinant pBR322 carrying foreign DNA at *Bam*HI site will not confer resistance to tetracycline, but are still resistant to ampicillin, which remains elsewhere.
- These recombinant cells are selected by replica plating method. The transformed cells are first plated on ampicillin containing medium and after the selection of transformed from non-transformed; the colonies are replica plated on medium containing tetracycline for screening of recombinant clones. After incubation, the viable colonies carrying pBR322 without DNA insert will appear and the positions in plate where the non-viable recombinant clones are present can be

easily identified. Using the original master plate, these recombinant clones are picked up and subcultured using the same procedure to obtain a pure recombinant clone.

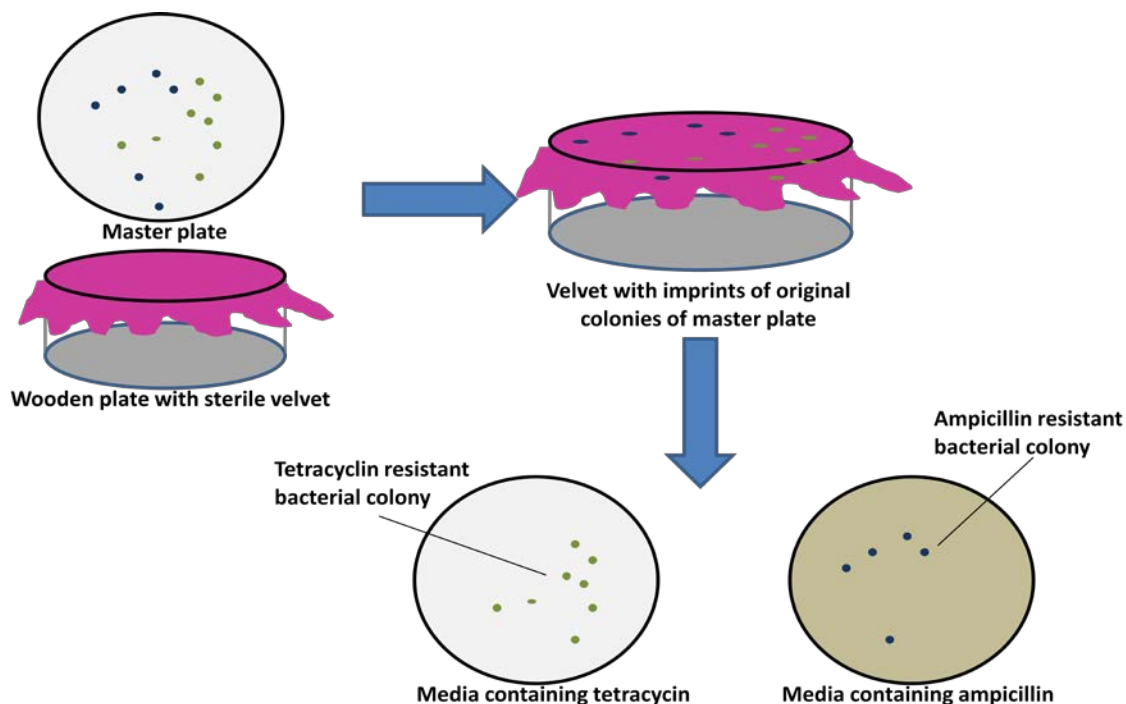


Fig 1-2.2.8.1: Selection through recombinant bacteria by replica plating

1-2.2.9 Screening for expression:

- Gene expression involves the synthesis of mRNA through transcription followed by synthesis of protein through translation. If the purpose of cloning is to express a foreign gene, it is necessary to check the expression both at the mRNA and protein level in terms of quality and quantity.
- Screening for foreign gene mRNA transcript can be done by:
 - ▶ **Northern blotting:** It involves the electrophoresis for the separation of RNA on the basis of size, and then transfer of RNA from the electrophoresis gel to the blotting membrane. It is then detected by a hybridization probe complementary to target sequence.

- ▶ **Reverse transcriptase PCR:** The RNA strand is reverse transcribed into its DNA complement (cDNA) using the enzyme reverse transcriptase and the resulting cDNA are then amplified using routine PCR method.
- Screening for foreign protein is done by:
 - ▶ **Western blotting:** It involves the process of electrophoretic separation of cellular proteins followed by transfer to a blotting membrane, incubation with a complimentary primary antibody probe and detection with a labelled secondary antibody.
 - ▶ **Activity/functional testing assay or staining:** The proteins which have specific functionality or activity can be tested in- vitro to confirm the presence of the same. For example, protease activity testing or staining.

Quantitative evaluation of the expression levels of the protein is necessary to achieve desired amount of protein.

1-2.3 Applications:

Molecular cloning serves as a tool for developing of various recombinant cells and organisms which has application both in basic and applied biological science as discussed below.

- Production of recombinant proteins: Genes encoding proteins with diagnostic, therapeutic or commercial value can be cloned, expressed and purified to obtain recombinant proteins in bulk with limited space and lesser time. Examples include Humulin – the human insulin expressed in *E. coli*.
- Gene therapy is used for correcting a disorder or deficiency.
- To study the structure and function of a particular gene using a model host organism.
- To study the regulation of gene expression during developmental stages.
- Complete genome sequence of an organism having large genome size can be facilitated by BAC or YAC library formation using rDNA technology.

Bibliography:

- Brown T.A. 2010. *Gene cloning and DNA analysis: an introduction* (6th edition); Willey Blackwell Ltd.
- Griffiths A.J.F, Miller J.H., Suzuki D.T, Lewontin R.C., Gelbert W.M. 2000. *An Introduction to Genetic Analysis* (7th edition); New York: W. H. Freeman.
- Robertis E.D.P.De, Robertis E.M.F. De. 2010. *Cell and Molecular biology* (8th edition); Lippincott Williams and Wilkins.

MODULE 1- LECTURE 3

TYPES, BIOLOGY AND SALIENT FEATURES OF VECTORS IN RECOMBINANT DNA TECHNOLOGY – PLASMID

1-3.1 Introduction:

DNA molecule used for carrying an exogenous DNA into a host organism and facilitates stable integration and replication inside the host system is termed as **Vector**. Molecular cloning involves series of sequential steps which includes restriction digestion of DNA fragments both target DNA and vector, ligation of the target DNA with the vector and introduction into a host organism for multiplication. Then the fragments resulted after digestion with restriction enzymes are ligated to other DNA molecules that serve as vectors.

In general, vectors should have following characteristics:

- Capable of replicating inside the host.
- Have compatible restriction site for insertion of DNA molecule (insert).
- Capable of autonomous replication inside the host (*ori* site).
- Smaller in size and able to incorporate larger insert size.
- Have a selectable marker for screening of recombinant organism.

1-3.2 Plasmids:

- Plasmids are naturally occurring extra chromosomal double-stranded circular DNA molecules which can autonomously replicate inside bacterial cells. Plasmids range in size from about 1.0 kb to over 250 kb.
- Plasmids encode only few proteins required for their own replication (replication proteins) and these proteins encoding genes are located very close to the *ori*. All the other proteins required for replication, e.g. DNA polymerases, DNA ligase, helicase, etc., are provided by the host cell. Thus, only a small region surrounding

the *ori* site is required for replication. Other parts of the plasmid can be deleted and foreign sequences can be added to the plasmid without compromising replication.

- The host range of a plasmid is determined by its *ori* region. Plasmids whose *ori* region is derived from plasmid Col E1 have a restricted host range. They only replicate in enteric bacteria, such as *E. coli*, *Salmonella*, etc. Plasmids of the RP4 type will replicate in most gram negative bacteria, to which they are readily transmitted by conjugation. Plasmids like RSF1010 are not conjugative but can be transformed into a wide range of gram -ve bacteria. Plasmids isolated from *Staphylococcus aureus* have a broad host range and can replicate in many other gram-positive bacteria.

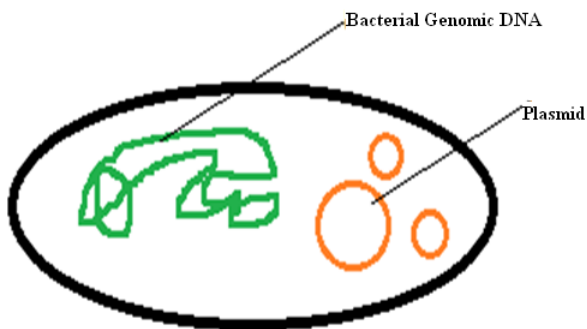


Fig 1-3.2: Bacterial Genomic DNA with plasmid

Some of the phenotypes which the naturally occurring plasmids confer on their host cells:

- Antibiotic resistance
- Antibiotic production
- Degradation of aromatic compounds
- Haemolysin production
- Sugar fermentation
- Enterotoxin production
- Heavy metal resistance
- Bacteriocin production
- Induction of plant tumors
- Hydrogen sulphide production

Most plasmids exist as double-stranded circular DNA molecules. However, the inter-conversion of super coiled, relaxed covalently closed circular DNA and open circular DNA is possible. Not all plasmids exist as circular molecules. Linear plasmids have been found in a variety of bacteria, e.g. *Streptomyces sp.* and *Borrelia burgdorferi*.

However, few types of plasmids are also able to replicate by integrating into bacterial chromosomal DNA; these are known as integrative plasmids or episomes. They are found mainly in prokaryotes but some eukaryotes are also found to harbour them. In prokaryotes they are found in *Escherichia coli*, *Pseudomonas* species, *Agrobacterium* species etc. In eukaryotes they are mainly found in *Saccharomyces cerevisiae*.

1-3.2.1 Types of Plasmids

The plasmids are divided into 6 major classes as described below depending on the phenotype:

- i) **Resistance or R plasmids** carry genes which give resistance to the bacteria from one or more chemical agents, such as antibacterial agents. R plasmids are very important in clinical microbiology as they can have profound consequences in the treatment of bacterial infections. Eg: RP4 plasmid, which is commonly found in *Pseudomonas* and in many other bacteria.
- ii) **Fertility or F plasmids** are conjugative plasmid found in F^+ bacterium with higher frequency of conjugation. F plasmid carries transfer gene (*tra*) and has the ability to form Conjugation Bridge (F pilus) with F^- bacterium. Eg: F plasmid of *E. coli*.
- iii) **Col plasmids** have genes that code for colicins, proteins that kill other bacteria. Eg: ColE1 of *E. coli*.
- iv) **Degradative plasmids** allow the host bacterium to metabolize unusual molecules such as toluene and salicylic acid. Eg TOL of *Pseudomonas putida*.

- v) **Virulence plasmids** confer pathogenicity on the host bacterium. Eg: Ti plasmids of *Agrobacterium tumefaciens*, which induce crown gall disease on dicotyledonous plants.
- vi) **Cryptic Plasmids** do not have any apparent effect on the phenotype of the cell harboring them. They just code for enzymes required for their replication and maintenance in the host cell.

Based on the origin or source of plasmids, they have been divided into two major classes: such as natural and artificial.

- i) **Natural plasmids**: They occur naturally in prokaryotes or eukaryotes. Example: ColE1.
- ii) **Artificial plasmids**: They are constructed *in-vitro* by re-combining selected segments of two or more other plasmids (natural or artificial). Example: pBR322.

1-3.3 Natural Plasmids

Few examples of naturally occurring plasmids and their characteristics are listed in table below

Plasmid	Size (kb)	Origin	Host range	Antibiotic resistance	Additional marker genes showing insertional inactivation
RSF1010	8.6	<i>E.coli</i> (strain K-12)	Broad host range	Streptomycin and sulfonamides.	None
ColE1	6.6	<i>E.coli</i>	Narrow host range	None	Immunity to colicin E1
R100	94.2	<i>E.coli</i>	<i>E.coli</i> K-12, <i>Shigella flexneri</i> 2b	Streptomycin, chloramphenicol, tetracycline	Mercuric (ion) reductase, ethidium bromide (EtBr) resistant protein.

Table 1-3.3: Characteristics of natural plasmids

1-3.3.1 RSF1010

- It is a naturally occurring plasmid isolated from *E.coli K-12*.
- This plasmid has broad host range in gram negative bacteria.
- The size of plasmid is 8694bp.
- Antibiotic resistance genes for Streptomycin and sulfonamides have present.
- The replication of RSF1010 starts either bi- or uni-directionally from unique *ori-V* region (2347-2742).
- It cannot initiate transformation independently but can be transferred to host bacterium in presence of helper plasmid.
- Genebank accession no. M28829.

1-3.3.2 ColE1:

- It is a naturally occurring multicopy plasmid obtained from *E.coli* (copy number is around 40).
- The size of this natural plasmid is 6646bp.
- It forms the basis of many artificial vectors used in molecular cloning.
- The natural ColE1 plasmid has genes for colicin E1 production. Colicin is an antibacterial toxin produced under stressed condition. Cells harboring the plasmid will have resistance against the toxin.
- For using in molecular cloning experiment, colicin genes are replaced with selection marker (antibiotic resistant) gene.
- Genebank accession no. M33100.

1-3.4 Artificial Plasmids:

Naturally occurring plasmids has several limitations; for example, some are stringent and not relaxed (pSC101), some has poor marker genes (ColE1), and some are too large (RSF2124). To overcome the limitations of natural vectors, artificial plasmid are designed by combining different elements from diverse sources.

Artificial plasmids vectors are classified into two broad types based on their use:

1. Cloning vector
2. Expression vector

Apart from the following, there is another class of vectors known as **shuttle vector**. Shuttle vectors can be propagated in two or more different host species (both in prokaryotes and eukaryotes). Hence, inserted DNA can be manipulated and replicated in two different cellular systems.

Cloning vectors are designed for efficient transfer of foreign DNA into the host. Expression vectors have efficient machinery for cloning and expression of foreign gene in the host system. Selection of a vector depends upon various criteria decided by the experimental goal.

1-3.5 Cloning Vector:

A cloning vector is defined as a vector used for replication of a cloned DNA fragment in a host cell. These vectors are frequently engineered to contain “ori” – origin of replication sites particular to the host organism. Examples of commonly used cloning vectors are: pUC18, pUC19, pBluescript vectors etc.

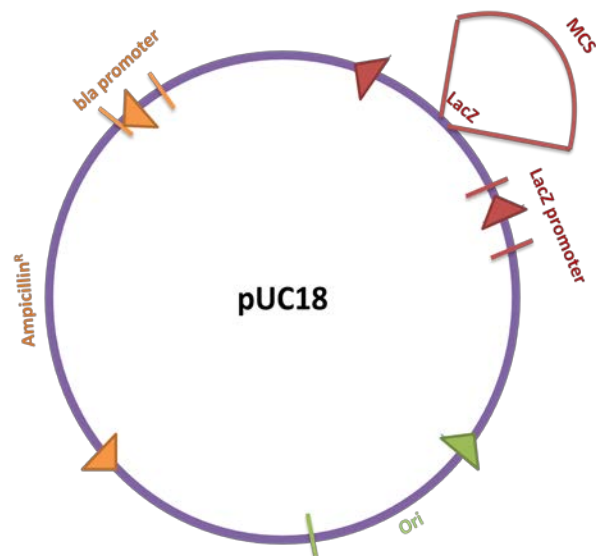


Fig 1-3.5: Cloning vector

Important features of a cloning vector used to carry DNA molecules are as follows:-

- **Stability in host cell:** Vectors should be stable in host cell after introduction and should not get lost in subsequent generations. This permits replication of vectors producing large copies of gene of interest.
- **Ability to control their own replication:** This property enables them to multiply and exist in high copy number.
- **Small size:** Ideal vector should be less than or equal to 10kb. The small size is essential for easy introduction in cell by transformation, transduction and electroporation.
- **Multiple cloning sites:** This property permits the insertion of gene of interest and plasmid re-circularization.
- **Should not be transferred by conjugation:** This property of vector molecule prevents recombinant DNA to escape to natural population of bacteria.
- **Selectable marker gene:** Vector molecules should have some detectable traits. These traits enable the transformed cells to be identified among the non-transformed ones. eg. antibiotic resistance gene.

1-3.6 Types of Cloning Vectors:

- Cloning vectors extensively used in molecular cloning experiments can be considered under following types: **plasmid**, **phage vector** and **cosmid**.
- Different vectors have different insert size and also vary in mode of replication inside the host.
- Mammalian genes are usually too large (~100 kb) and thus suffer from restrictions in complete inclusion with the conventional cloning vectors having limited insert size.
- Vectors engineered more recently, known as artificial chromosomes, have overcome this problem by mimicking the properties of host cell chromosomes. They have much larger insert size than other vectors.

Vector	Insert size	Source	Application
Plasmid	≤ 15 kb	Bacteria	Subcloning and downstream manipulation, cDNA cloning and expression assays
Phage	5-20 kb	Bacteriophage λ	Genomic DNA cloning, cDNA cloning and expression library
Cosmid	35-45 kb	Plasmid containing a bacteriophage λ <i>cos</i> site	Genomic library construction
BAC (bacterial artificial chromosome)	75-300 kb	Plasmid containing <i>ori</i> from <i>E.coli</i> F- plasmid	Analysis of large genomes
YAC (yeast artificial chromosome)	100-1000 kb (1 Mb)	<i>Saccharomyces cerevisiae</i> centromere, telomere and autonomously replicating sequence	Analysis of large genome, YAC transgenic mice
MAC (mammalian artificial chromosome)	100 kb to > 1 Mb	Mammalian centromere, telomere and origin of replication	Under development for use in animal biotechnology and human gene therapy

Table 1-3.6: Different type of cloning vectors

1-3.7.1 Examples of Cloning Vector:

1-3.7.1.a pBR322

- pBR322 is a widely-used *E. coli* cloning vector. It was created in 1977 in the laboratory of Herbert Boyer at the University of California San Francisco. The *p* stands for "**plasmid**" and *BR* for "**Bolivar**" and "**Rodriguez**", researchers who constructed it.
- pBR322 is 4361 base pairs in length.
- pBR322 plasmid has the following elements:
 - ▶ “*rep*” replicon from plasmid pMB1 which is responsible for replication of the plasmid.
 - ▶ “*rop*” gene encoding Rop protein. Rop proteins are associated with stability of RNAI-RNAII complex and also decrease copy number. The source of “*rop*” gene is pMB1 plasmid.
 - ▶ “*tet*” gene encoding tetracycline resistance derived from pSC101 plasmid.
 - ▶ “*bla*” gene encoding β lactamase which provide ampicillin resistance (source: transposon Tn3).

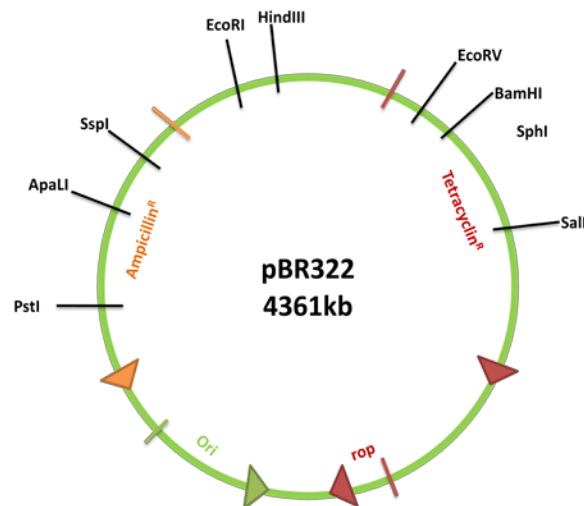


Fig 1-3.7.1.A: Plasmid pBR322.

1-3.7.1.b pUC plasmids:

- pUC plasmids are small, high copy number plasmids of size 2686bp.
- This series of cloning vectors were developed by Messing and co-workers in the University of California. The p in its name stands for plasmid and UC represents the University of California.
- pUC vectors contain a *lacZ* sequence and multiple cloning site (MCS) within *lacZ*. This helps in use of broad spectrum of restriction endonucleases and permits rapid visual detection of an insert.
- pUC18 and pUC19 vectors are identical apart from the fact that the MCS is arranged in opposite orientation.
- pUC vectors consists of following elements:
 - ▶ pMB1 “*rep*” replicon region derived from plasmid pBR322 with single point mutation (to increase copy number).
 - ▶ “*bla*” gene encoding β lactamase which provide ampicillin resistance which is derived from pBR322. This site is different from pBR322 by two point mutations.
 - ▶ *E.coli* lac operon system.
- “*rop*” gene is removed from this vector which leads to an increase in copy number.

An MCS is a short DNA sequence consisting of restriction sites for many different restriction endonucleases. MCS escalates the number of potential cloning strategies available by extending the range of enzymes that can be used to generate a restriction fragment suitable for cloning. By combining them within a MCS, the sites are made contiguous, so that any two sites within it can be cleaved simultaneously without excising vector sequences. The MCS is inserted into the *lacZ* sequence, which encodes the promoter and the α -peptide of β -galactosidase. Insertion of the MCS into the *lacZ* fragment does not affect the ability of the α -peptide to mediate complementation, while cloning DNA fragments into the MCS does. Therefore, recombinants can be detected by blue/white screening on growth medium containing X gal in presence of IPTG as an inducer.

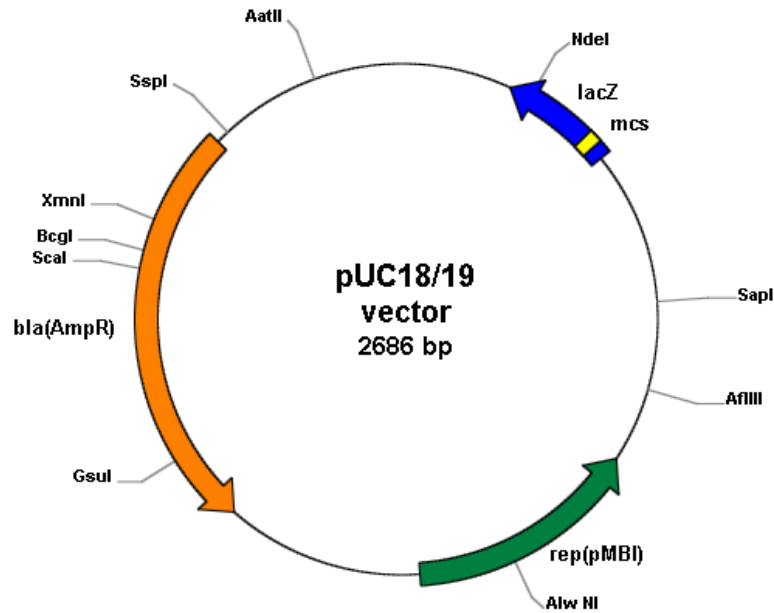


Fig 1-3.7.1.B: pUC plasmid

1-3.8 Expression Vector:

A vector used for expression of a cloned DNA fragment in a host cell is called as an expression vector. These vectors are frequently engineered to contain regulatory sequences that act as promoter and/or enhancer regions and lead to efficient transcription of the insert gene. Commonly used expression vector series are: pET vectors, pBAD vectors etc.

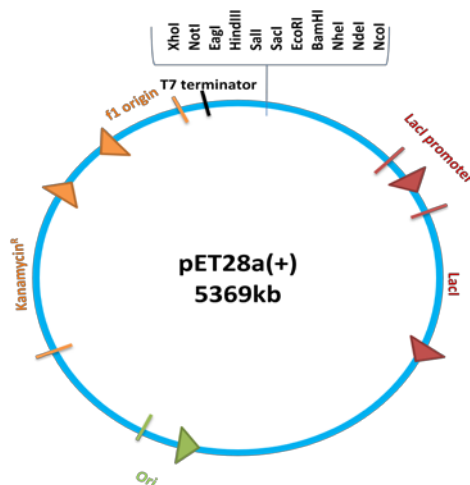


Fig 1-3.8: Expression vector

For an expression vector following features are essential:

- Promoter: Promoter is a sequence which is recognized by sigma subunit of RNA polymerase which is required for initiation of transcription of gene of interest.
- Terminator: It is a DNA element present at the end of a gene where transcription of gene ends. Terminator is short nucleotide sequences which can base pair with itself to form hair pin loop.
- Ribosome binding site: It is a short nucleotide sequence recognized by the ribosome as the point at which it should attach to the messenger molecule. The initiation codon of the gene is always a few nucleotides downstream of this.

1-3.8.2 Examples of Expression Vector:

1-3.8.2.a pET vector:

- pET vector system is a cloning and expression vector system for recombinant protein production in *E.coli*. This product is registered under trademark of Novagen Inc.
- The original pET vector system was constructed by Studier and colleagues. That plasmid is developed at Novagen with enhanced characteristics.
- Target genes are cloned under strong T7 bacteriophage promoter.
- The expression of the target protein is inducible by providing T7 RNA polymerase in the host cell as an inducing signal.
- Target gene is initially cloned to host cell that do not contain T7 RNA Polymerase, thus increasing plasmid stability.
- Once stabilized in a non-expression host, the recombinant plasmid is transferred to a expression host having T7 RNA Polymerase gene in the genome.
- Ampicillin and kannamycin resistance genes are available in pET vectors as selection marker.
- pET28 and pET32 are the most commonly used pET vectors.

1-3.9 Specialized Expression Vectors:

In molecular biology, vectors are generally designed for cloning a foreign gene into a host genome so that the host produces proteins which are normally not produced by host. But, apart from these applications, different specialized vectors have been constructed to achieve different application in genetic and molecular biology studies. The vectors constructed for thus specialized functions are termed as specialized vectors. Molecular and genetics study of a gene or protein can be aided by specialized vectors. Some of the applications of specialized vectors have been discussed below-

1-3.8.1: Promoter Probe Vectors:

Specialized vectors used for identification of efficient promoter region in a DNA segment are termed promoter probe vectors. Promoter-less reported genes (*lacZ*, GFP etc) are used for construction of promoter probe vectors. The expression of the reporter genes can be monitored and quantified easily using various biochemical or fluorescent techniques. Fusion of DNA fragment containing a promoter region upstream of the reporter gene drives the expression of the reported gene. However, there is no guarantee that the DNA sequence that behaves as promoter in recombinant host can behave in the same way in its native host (Pseudo- promoter). Further characterization is necessary to define a true novel promoter. Some of the widely used promoter probe vectors families are: pOT (eg. pRU1161, pRU1097 etc) and pJP2 (eg. pRU1156, pRU1157 etc). pOT vectors have higher copy number but lower stability as compared to pJP2 vectors.

1-3.9.2 Gene Fusion Vectors:

Fusion of one gene to another gene in order to produce a fusion protein is widely used in molecular biology studies. Fusion proteins are generated by cloning two or more target genes with a reporter gene (*His-tag*, *gfp*, *rfp*, *lacZ* etc) by using gene fusion vectors. Fusion proteins may provide improved properties like easy isolation and purification of target protein (*His-tag*), easy monitoring of gene expression level (GFP, RFP, *lacZ*), intracellular protein localization studies (GFP, RFP, LUC) etc. The target gene is cloned downstream of the promoter region present in the vector. Depending on the requirement, the target protein can be cloned either to the N-terminal or C-terminal

of the reporter protein. Different vectors have been commercially available to provide such flexibility in cloning site and reporter gene.

1-3.10 Viral Vectors:

In recombinant molecular biology, virus particles has been modified to use as a carrier of nucleic acid into a cell, termed as viral vectors. Viral vectors are highly efficient in transferring target DNA/RNA segment to the host cells with high specificity. Wild type virus are modified by deleting the non-essential genes and incorporating exogenous nucleic acid segments to construct a viral vector. Viral vectors have wide application in gene therapy and targeted drug delivery systems. Main advantages of viral vectors are-high transfer efficiency and high cell specificity. Although, there are certain safety issues associated with viral vectors and careful handling is essential during the experimental procedure.

Commonly used viral vectors are- Adenovirus, retrovirus, lentivirus, adeno associated virus (AAV), herpes simplex virus (HSV) etc. The properties of different viral vectors are summarized in the table below.

Viral vector	Insert type	Insert size	Immunogenicity	Host genome integration
Adeno virus	DNA	2-8 kb	Very high	Non integrating
Retro virus	RNA	2-8 kb	Low	Integrating
Lentivirus	RNA	7-18 kb	Low	Integrating
Adeno associated virus (AAV)	DNA	4.5 kb	Low	Non integrating
Herpes simplex virus (HSV)	DNA	>30kb	Low	Non integrating

Table 1-3.10: Types of viral vectors

1-3.10.1 Simian Virus 40 (SV40):

Simian virus 40 (SV40) was the first mammalian expression vector whose genome size is 5.2kb. It is a DNA virus which can infect human as well other mammalian cell lines. SV40 may integrate into the host genome, permitting stable transmission of insert DNA to daughter cells. Recombinant SV40 vectors (rSV40) display some unique features:

- SV40 is a well-known virus and nonreplicative vectors are easy-to-make
- SV40 can be produced in high titers (10^{12} IU/ml).
- They can infect both resting and dividing cells.
- Stable transgene expression can be achieved in a wide range of cell types.

The major disadvantage of SV40 vector is the low packaging capacity with insert size of <5kb.

1-3.10.2 Baculoviral Vectors:

Baculovirus is a DNA virus with host range restricted to invertebrates, mostly insects. The baculovirus expression system has been used extensively for the expression of recombinant proteins in insect cells. Baculovirus is a group of insect virus and *Autographa californica* nucleopolyhedrovirus (AcMNPV) is the most extensively studied virus under this family.

The infection of AcMNPV is initiated by replication and transcription of the DNA genome inside the nucleus and the assembly of the nucleocapsids. The nucleocapsids then bud off from the plasma membrane and initiate systemic infection.

Although, baculoviral vectors can transfect only insect cells, recombinant baculoviral vectors have been constructed containing mammalian cell specific promoters which can be used to infect mammalian cells as well.

Advantages of baculoviral vectors:

- Since insects cells are high eukaryotes, desired post-translational modification of complex protein can be achieved.
- They have higher packaging capacity of insert.
- Lower biosafety issue.
- High level of protein expression.

There are few drawbacks of baculoviral vectors such as-

- Foreign protein expression using an insect system is more complex and time consuming than a bacterial system.
- Sometime protein post-translational processing may be sub-optimal to compensate the secretory pathway of the protein.

1-3.11 Yeast Vector System:

Cloning and expression of a gene using yeast system has several advantages over E.coli system due to presence of eukaryotic post-translational modification machinery. Expression of complex proteins with proper modification and folding can be achieved by yeast eukaryotic system. These vectors have yeast origin of replication (ARS) for replication and maintenance in the yeast system and bacterial ori for maintaining inside a bacterial system.

Different types of yeast vector include YIp (yeast integrative plasmid), YE_p (yeast episomal plasmid), YRp (yeast replicating plasmid), YC_p (yeast centromere plasmid) etc.

YIp (yeast integrative plasmid) can integrate to the host genome by homologous recombination. This generally yields a single copy of recombinant vector DNA integrated to the host genome.

YE_p (yeast episomal plasmid) can be maintained in the yeast system as an autonomously replicating episomal plasmid. This vector contains a part of 2 μ plasmid which is essential for autonomous replication of the vector inside the yeast.

YRp (yeast replicating plasmid) are used to obtain a high copy number inside the host (upto 100 copy number).

YCp (yeast centromere plasmid) possess a centromeric region in addition to ARS which facilitates the mitotic segregation of the linear plasmid during replication. The copy number of this vector is essentially one per cell.

Bibliography:

- Cohen S.N., Chang A.C., Boyer H.W., Helling R.B. 1973. Construction of biologically functional bacterial plasmids *in vitro*. *Proc Natl Acad Sci USA* 70: 3240–3244.
- Esposito D, Garvey L.A, Chakiath C.S. 2009. Gateway Cloning for Protein Expression. *Met Mol Biol* 498: 31-54.
- Karunakaran R, Mauchline T.H, Hosie A.H, Poole P.S. 2005. A family of promoter probe vectors incorporating autofluorescent and chromogenic reporter proteins for studying gene expression in Gram-negative bacteria; *Microbiology*.151(10):3249-3256.
- Kay M. A, Glorioso J. C., Naldini L. 2001. Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics; *Nat. Med. Rev.* 7(1): 33-40.
- Kost T. A. and Condreay J. P. 2002. Recombinant baculoviruses as mammalian cell gene-delivery vectors; *Trends in Biotechnol* 20(4).
- Kroll J, Klintner S, Schneider C, Voss I, Steinbüchel A. 2010. Plasmid addiction systems: perspectives and applications in biotechnology. *Microb Biotechnol.* 3(6):634-57.
- Nilsson B., Abrahmsen L., Uhlen M. 1985. Immobilization and purification of enzymes with Staphylococcal protein A gene fusion vectors; *The EMBO Journal* 4(4):1075-1080.
- pET System Manual (10th edition); Novagen Inc.
- Sikorski R. S., Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*; *Genetics* 122:19-27.

- Yu-chen Hu. 2005. Baculovirus as a highly efficient expression vector in insect and mammalian cells; *Acta Pharmacologica Sinica* 26 (4): 405–416.
- Ong Y, Ward J, Nagl S. 2008; Genome Database of Naturally Occurring Plasmids; URL: <http://www.biochem.ucl.ac.uk/bsm/PLASMID/mainpage.htm>
- Honda Y, Sakai H, Hiasa H, Tanaka K, Komano T, Bagdasarian M. 1991. Functional division and reconstruction of a plasmid replication origin: Molecular dissection of the *oriV* of the broad-host-range plasmid RSF1010; *Proc. Natl. Acad. Sci. USA*. 88:179-183.

MODULE 1- LECTURE 4

TYPES, BIOLOGY AND SALIENT FEATURES OF VECTORS IN RECOMBINANT DNA TECHNOLOGY- PHAGES, COSMIDS, FOSMIDS, PHAGMIDS AND ARTIFICIAL CHROMOSOMES

1-4.1 Introduction:

You now know that in recombinant DNA technology, vectors are used as carrier of foreign DNA into the host organism. Apart from bacterial plasmids, several other modified vectors are constructed using molecular tools. These “hybrid” vectors are designed by combining different components from various origins (bacteriophage, F plasmid etc) to create a capacity to load larger insert size and higher transfection efficiency.

1-4.2 Phage Vectors:

To insert DNA fragments of more than 10 kb, normally plasmids are not the suitable vehicles, as large inserts may trigger plasmid rearrangement or affect plasmid replication. This leads to development of a new class of vectors based on bacteriophages. Amongst various bacteriophages available such as λ , T4, T5, and T7 phages; the λ phage gained favourable attention due to its unique life cycle.

λ phage

Bacteriophage λ contains ~49kb of DNA and has a very efficient mechanism for delivering its genome into a bacterium. Two key features contribute to its utility as a vector to clone larger DNA fragments:

1. One-third λ genome is nonessential and could be replaced with foreign DNA. Approximately 24.6kb of λ genome can be deleted, hence maximum insert size could be upto 26 kb.

2. Packing of DNA in phage could only take place if the size is between 40 and 52 kb long, a constraint that can be used to ensure packaging.

Two problems had to be addressed before λ -based cloning vectors could be developed:

- The size limitation of the insert is determined by the genome size of phage λ (distance between the cos sites). The size range of the modified genome size should be within the range between 78-105% of the genome size for proper packaging. If >2.4kb is inserted to full length λ vector, the packaging efficiency is reduced. Hence λ vector should be smaller in size than wild type λ genome.
- The large λ genome has a few unique recognition sequences for bacterial restriction endonuclease. Bacterial restriction digestion system may target the modified vector and cleave the λ DNA molecule. This limitation can be overcome by replacing or mutating the restriction sites.

Two types of vector have been developed using λ genome:

1) Insertion vectors:

- Foreign DNA sequence is inserted into the λ genome without any significant change of the wild type genome.
- Smaller insert size (upto ~10kb).
- They may contain a multiple cloning site inserted in *lacZ* system for screening of recombinant bacterial colonies.
- Can be used to clone smaller DNA molecule.
- Eg: λ ZAP, λ gt etc.

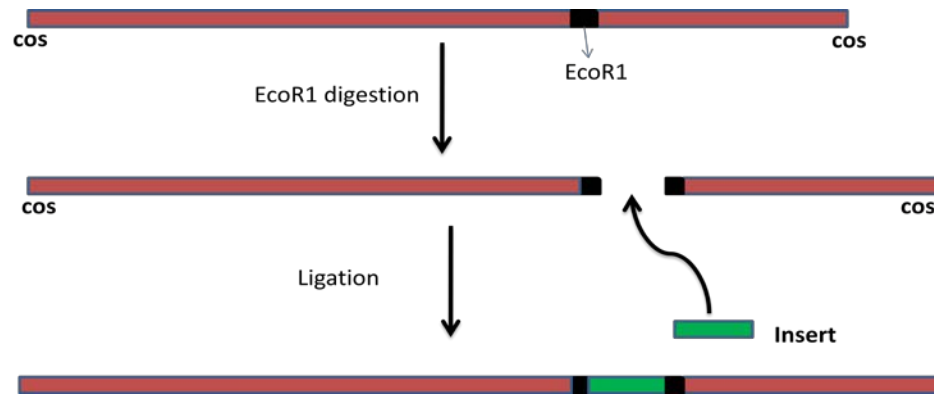


Fig 1-4.2.1: Insertion vector

2) Replacement vectors:

- Full length λ molecule having two identical restriction sites flanked by “stuffer fragment”.
- Stuffer fragment is replaced by foreign DNA during restriction cloning.
- The vector without the foreign insert cannot be packaged due to the size limitation (smaller than the required).
- Insert size ranges between 10-23 kb.
- Example: λ EMBL 3, λ EMBL 4, λ DASH etc.

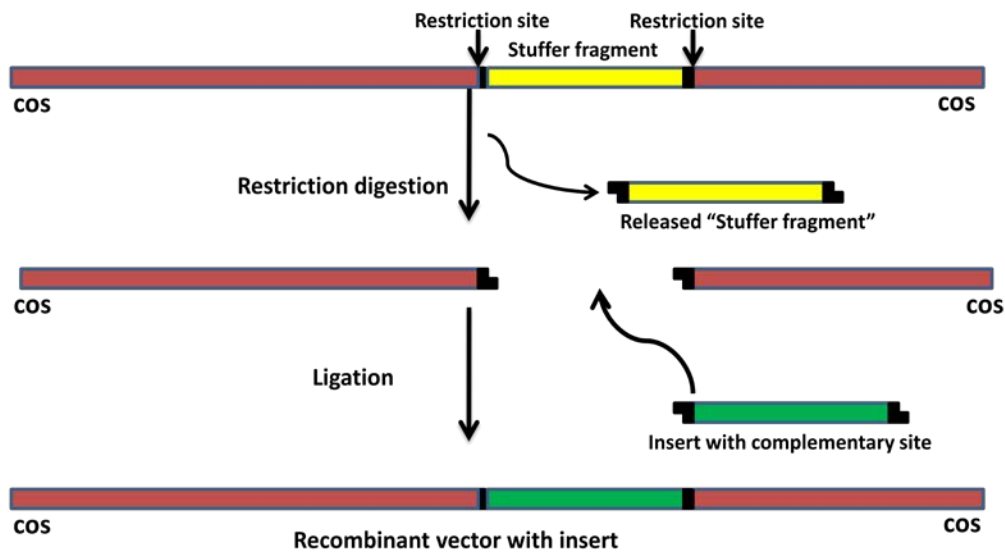


Fig 1-4.2.2: Replacement vector

1-4.2.1 Features of λ Phage Vectors:

The λ genome is linear, but in the ends has 12-nucleotides overhangs, termed as *cos* sites, which are complementary to each other. A λ cloning vector can be circularized using *cos* site which can be manipulated and replicated inside *E. coli* via the process of transfection.

Alternatively, a more efficient uptake system called *in vitro* packaging can be utilized. Treatment with the appropriate restriction endonuclease followed by ligation in presence of insert DNA, produces ‘left arm-new DNA-right arm’ concatemers that are then added to an *in vitro* packaging mix to form λ phage particles. These phages are then co-incubated with *E. coli* cells, and the infection process naturally transports the vector plus new DNA into the bacteria. Bacteria that are infected with the packaged cloning vector die within about 20 minutes and several rounds of phage replication and bacterial lysis forms a zone of clearing, called a plaque.

1-4.2.2 M13 Phage Vectors:

M13 phage is filamentous phage that infects *E. coli* via F-pilus. The genome is a single stranded circular DNA of size ~6.4kb surrounded by a proteinaceous coat. The DNA strand present in phage is called plus (+) strand. After entering to *E. coli* host, it converts into double stranded DNA molecule called **replicative form (RF)** by utilizing bacterial machinery. M13 phage as cloning vector can be obtained in both single stranded as well as double stranded form. Replicative form double stranded vector are modified and replicated inside *E. coli* host similar to a plasmid vector. Single stranded vectors can be isolated by collecting M13 phage. M13 vectors have useful application in following areas:

- DNA sequencing
- Mutagenesis study
- probe generation
- Phage display

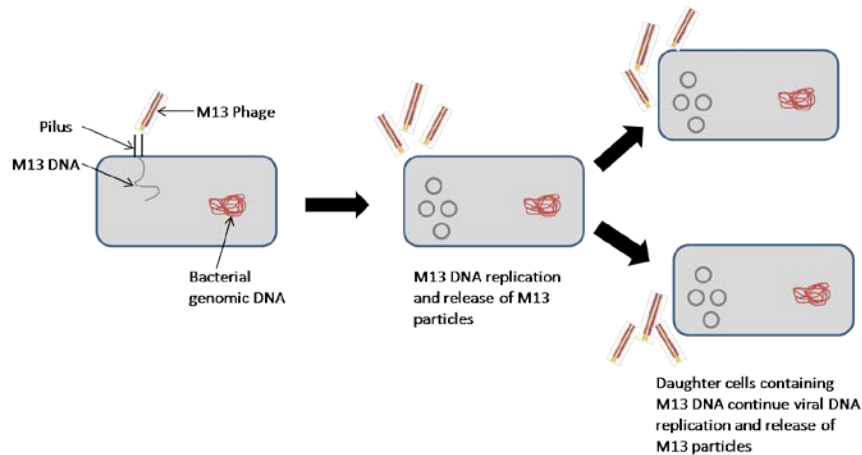


Fig1-4.2.2: M13 Phage Vectors

1-4.3 Cosmid:

- A cosmid, first described by **Collins** and **Hohn** in 1978, is a type of hybrid plasmid with a bacterial “ori” sequence and a “cos” sequences derived from the lambda phage.
- *Cos* site is the sequence required by a DNA molecule in order to be recognized as a ‘ λ genome’ by the proteins that package DNA into λ phage particles.
- Cosmid DNA containing particles are as transmittable as real λ phages, but once inside the cell, the cosmid cannot control synthesis of new phage particles and instead replicates as a plasmid.
- Recombinant DNA is therefore obtained from colonies rather than plaques. They frequently also contain a gene for selection such as antibiotic resistance.
- They are able to load 37 to 52 kb of DNA, while normal plasmids are able to carry only 1–26 kb.
- Sometimes helper phage is used to assist in packaging of cosmid inside phage. Helper phage provides the essential proteins required for packaging which are lacked by cosmid vector.
- For packaging into a phage, concatemer formation is required (cosmid-insert-cosmid). This is generated by using two *cos* sites flanked by the insertion site for foreign DNA. Providing the inserted DNA in the right size, *in vitro* packaging cleaves the *cos* sites and replaces the recombinant cosmids in mature phage particles.

- Recombinant λ phages are used to infect an *E. coli* culture. Infected cells are plated on a selective medium and antibiotic-resistant colonies are grown. All colonies are recombinants, as non-recombinant linear cosmids are too small to be packaged into λ heads.
- Cosmids are widely exploited to build genomic libraries. The upper limit for the length of the cloned DNA is set by the space available within the λ phage particle. New DNA insert of size up to 44 kb can be inserted before the packaging limit of the λ phage particle is reached.

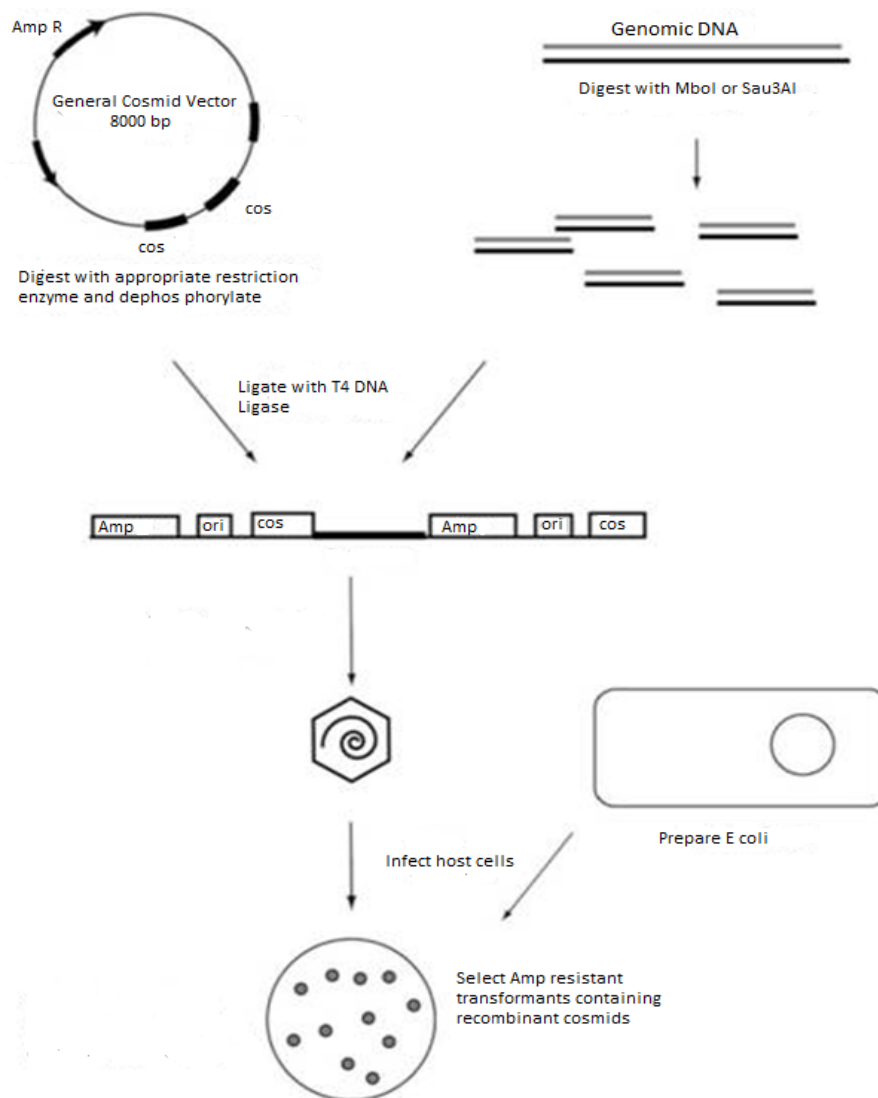


Fig 1-4.3: Schematics for Cosmid Library Construction

1-4.3.1 Limitation of Cosmid vector:

- Slower replication
- Higher frequency of recombination inside bacterial host.
- Unstable inside *E.coli* host and thus easy to lose vector.

Example:

pJB8 is 5.4 kb in size and carries the ampicillin-resistance gene (amp^R), a segment of λ DNA containing the *cos* site, and an *Escherichia coli* origin of replication (*ori*).

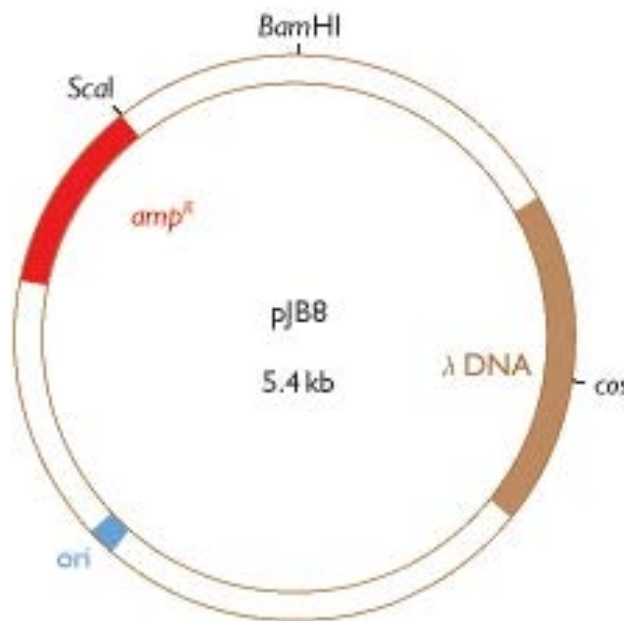


Fig 1-4.3.1: Schematic representation of pJB8

1-4.4 Fosmid:

- Fosmids are similar to cosmids, however they are primarily based on bacterial F-plasmid.
- **Simon and co-workers**, in the year 1992, first developed F-factor based vector named as **pFOS** for stable propagation of cosmid sized human genomic DNA inserts.
- They carry the F plasmid origin of replication and a λ *cos* site.
- Fosmids can carry up to 40 kb of insert DNA.

- The cloning vector is limited, as a host (usually *E. coli*) can only contain one fosmid molecule. Low copy number offers higher stability as compared to high copy number cosmids.
- Fosmids have high structural stability and have been found to maintain human DNA effectively even after 100 generations of growth. It is ideal to use a fosmid vectors for constructing genomic and meta-genomic libraries.

Fosmids contain several functional elements as discussed below,

- ▶ OriT (Origin of Transfer): The sequence which marks the starting point of conjugative transfer.
- ▶ OriV (Origin of Replication): The sequence starting with which the plasmid-DNA will be replicated in the recipient cell.
- ▶ *tra*-region (transfer genes): Genes coding the F-Pilus and DNA transfer process.
- IS (Insertion Elements): so-called "selfish genes" (sequence fragments which can integrate copies of themselves at different locations).

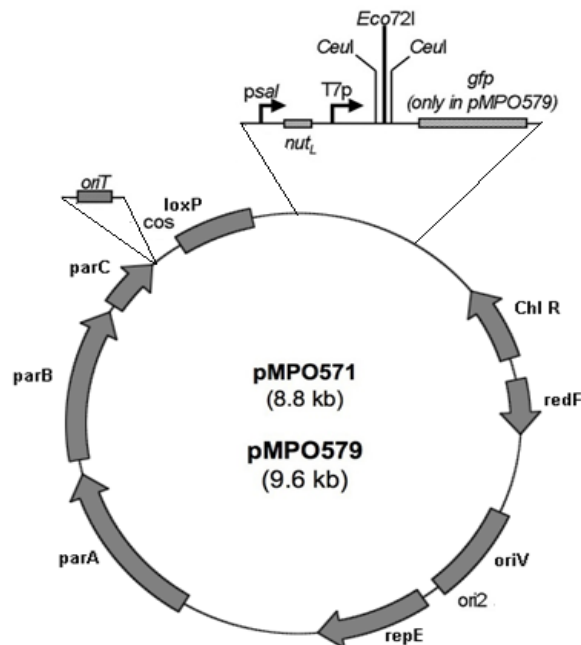


Fig 1-4.4: Schematic diagram of fosmids derived from pCC1FOS-CeuI

1-4.5 Phagemid

Although M13 vectors are very useful for the production of single-stranded recombinant genes, they have certain disadvantages. There is a limit to the size of DNA fragment that can be cloned in an M13 vector, with 1.5 kb being the ideal capacity, although fragments up to 3 kb have occasionally been cloned. To overcome this limitation, phagemid vectors were developed by combining a part of the M13 genome with plasmid DNA.

- Phagemids are cloning vectors developed as a hybrid of the filamentous phage M13 and plasmids to produce a vector that can get packed as a phage particle but also can propagate as a plasmid.
- They contain an origin of replication (ori) for double stranded replication inside *E. coli* host, as well as an “f1 ori” to enable single stranded replication and packaging into phage particles. Many commonly used plasmids contain an f1 ori and are thus phagemids.
- Phagemid generally encode no or only one of the capsid proteins of virus. Other structural and functional proteins necessary for phage lifecycle are provided by the helper phage.
- The components present in a phagemid vector are:
 - ▶ Origin of replication (ori) of a plasmid.
 - ▶ Intergenic region (IG region) which contains the packaging signal for the phage particle and also has replication origin inside phage.
 - ▶ A gene encoding phage coat protein.
 - ▶ A selection marker.
 - ▶ Restriction enzyme recognition sites.

Phagemid vectors are commonly used for “phage display technology” by which a broad range of proteins and peptides can be expressed as fusions to phage coat proteins and displayed on the viral surface. The advantage of phagemid vectors is that double stranded phagemid vectors can be converted into single stranded vectors and packaged into virion particles by infecting the cells with helper phage.

- Phagemid has certain advantages over phage vectors:
 - ▶ The carrying capacity of phagemid is higher than phage vectors.
 - ▶ Phagemid has higher efficiency in transformation than phage vectors.
 - ▶ Phagemids are genetically more stable than recombinant phage vectors.
 - ▶ Phagemids can be exploited to generate single stranded DNA template for sequencing purposes.
 - ▶ Single stranded phagemid vectors inside the phage can be targeted for site-directed mutagenesis.
 - ▶ Single stranded vectors can be used to generate hybridization probes for mRNA or cDNA.

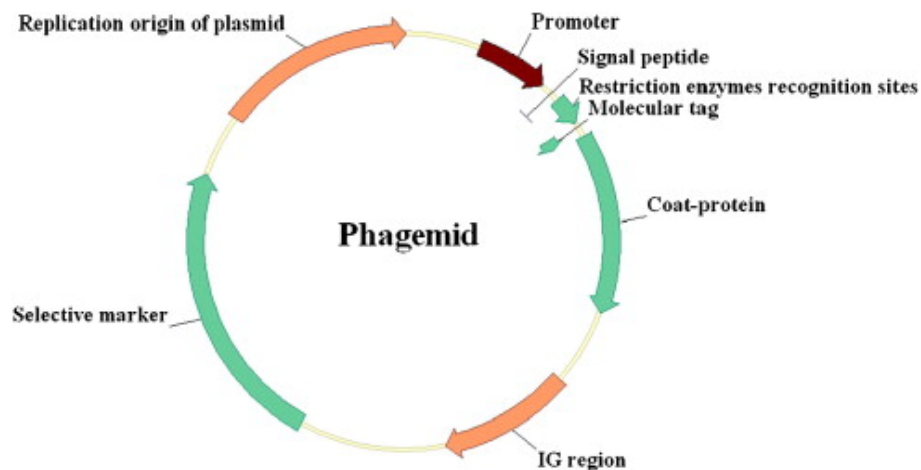


Fig1-4.5: Phagemid vector for phage display

[**Source:** Huan Qi, Haiqin Lu, Hua-Ji Qiu, Valery Petrenko and Aihua Liu (2012); Phagemid Vectors for Phage Display: Properties, Characteristics and Construction; *J.Mol. Bio* (417), 129-143]

1-4.5.1 Examples of Phagemid:

1-4.5.1.A pEMBL

One of the first hybrid phagemid vectors was pEMBL constructed in 1983. They are characterized by the presence of –

- 1) The *bla* gene as selectable marker for ampicillin resistance.
- 2) A short segment coding for the alpha-peptide of beta-galactosidase (*lacZ*) and containing a MCS.
- 3) The intragenic (IG) region of phage F1.

These vectors have been used successfully for DNA sequencing with the dideoxy method, and can be used for other purposes for which M13 derivatives are used. However, the pEMBL plasmids have the advantage of being smaller than M13 vectors, and the purification of DNA is simpler. In addition, long inserts have a higher stability in pEMBL plasmids than M13 vectors.

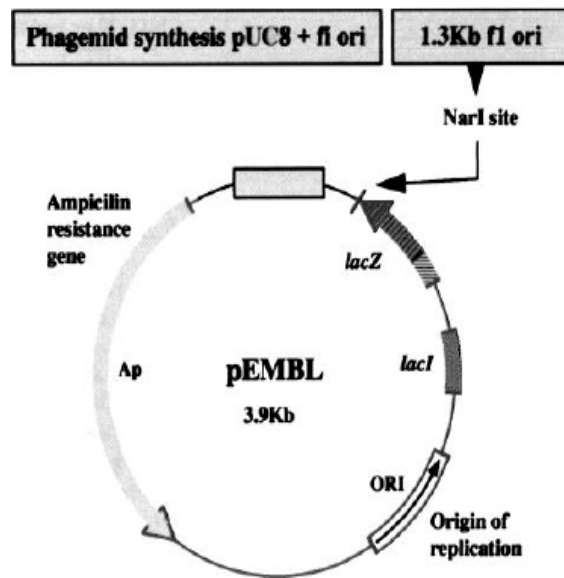


Fig: 1-4.5.1.A pEMBL vector

1-4.6 Artificial Chromosomes:

Artificial chromosomes are DNA molecules assembled *in vitro* from defined constituents that can function like natural chromosomes.

Types of artificial chromosomes:

- i) BACs: Bacterial artificial chromosomes
- ii) YACs: Yeast artificial chromosomes
- iii) MACs: Mammalian artificial chromosomes
- iv) HACs: Human artificial chromosomes
- v) PACs: P1-derived artificial chromosomes

Some of the artificial chromosomes are discussed below-

1-4.6.1 Bacterial Artificial Chromosomes: BAC:

Bacterial artificial chromosomes (BACs) are designed for the cloning of large DNA insert (typically 100 to 300 kb) in *E. coli* host. BAC vectors contain a single copy F-plasmid origin of replication (ori).

The F (fertility) plasmid is relatively large and vectors derived from it have a higher capacity than normal plasmid vectors. F-plasmid has F (fertility) factor which controls the replication and maintain low copy number. Also conjugation can take place between F⁺ bacteria (male) and F⁻ bacteria (female) to transfer F-plasmid via pilus.

Common gene components of a bacterial artificial chromosome are:

- 1) **oriS**, **repE** – F for plasmid replication and regulation of copy number.
- 2) **parA** and **parB** for maintaining low copy number and avoiding two F plasmids in a single cell during cell division.
- 3) A selectable marker for antibiotic resistance; some BACs also have *lacZ* at the cloning site for blue/white selection.
- 4) T7 and Sp6 phage promoters for transcription of inserted genes.

The *par* genes, derived from F plasmid assist in the even distribution of plasmids to daughter cells during cell division and increase the likelihood of each daughter cell carrying one copy of the plasmid, even when few copies are present. The low number of copies is useful in cloning large fragments of DNA because it limits the opportunities for unwanted recombination reactions that can unpredictably alter large cloned DNA over time.

The first BAC vector, **pBAC108L**, did not contain a selectable marker for recombinants. Thus, positive recombinants had to be identified by colony hybridization. Two widely used BAC vectors, **pBeloBAC11** and **pECBAC1**, are derivatives of pBAC108L in which the original cloning site is replaced with a *lacZ* gene carrying a multiple cloning site. **pBeloBAC11** has two *EcoRI* sites, one in the *lacZ* gene and one in the CMR gene, whereas pECBAC1 has only one *EcoRI* site in the *lacZ* gene. Further improvements to BACs have been made by replacing the *lacZ* gene with the *sacB* gene which is a negative selection marker. The product of *sacB* gene is levansucrase which can convert sucrose present in the media into levan, a toxin for the bacteria. Hence the colonies without insert would have intact *sacB* gene and thus cells die before forming colonies.

The F plasmid is relatively large and vectors constructed on it have a higher capacity for accepting inserted DNA. A similar cloning vector called a P1-derived artificial chromosome or PAC has also been produced from the bacterial P1 bacteriophage DNA. Both BACs and PACs can be used to clone fragments of 300kb and longer. They are often used to sequence the genome of organisms in genome projects.

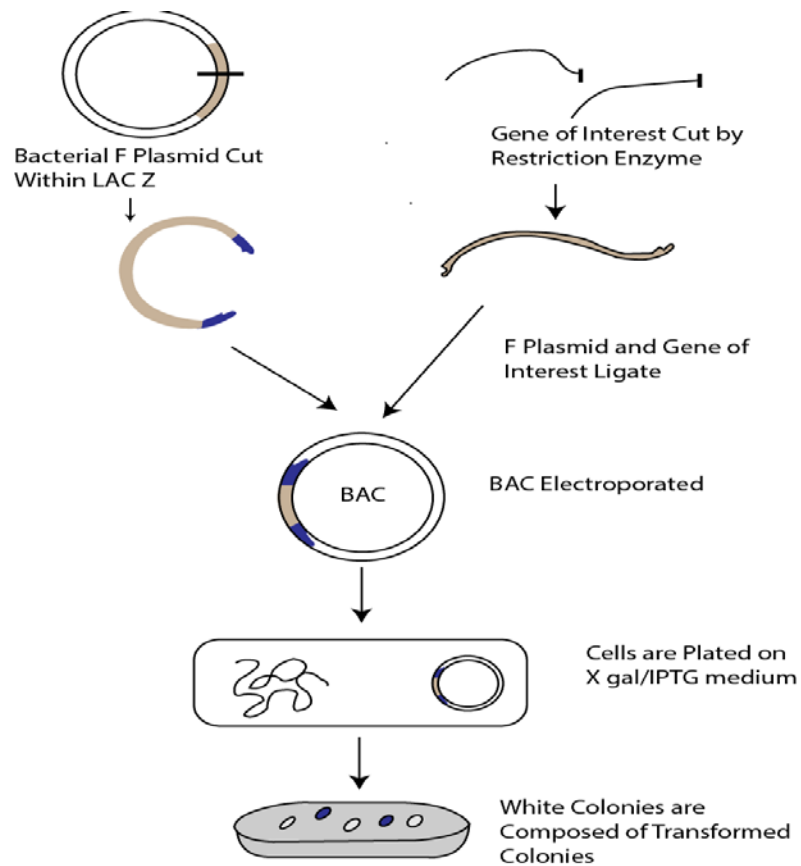


Fig 1-4.6.1: Transforming a Bacterium Using a BAC Vector

1-4.6.2 Yeast Artificial Chromosomes: YAC

- First described in 1983 by **Murray** and **Szostak**, a yeast artificial chromosome has sequences to exist inside *E. coli* as a circular plasmid and contains sequences to maintain as linear nuclear chromosome in yeast.
- As YAC vectors can accommodate 100-500 kb of insert DNA. The number of clones in a genomic library can be greatly reduced.
- YAC vectors have following elements:
 - ▶ *E. coli* origin of replication
 - ▶ Yeast origin of replication
 - ▶ Elements of eukaryotic yeast chromosome (centromere and telomere region)
 - ▶ Selection markers for both the host.
- YAC vector is initially propagated as circular plasmid inside bacterial host utilizing bacterial *ori* sequence. Circular plasmid is cut at specific site using restriction enzymes to generate a linear chromosome with two telomere sites at terminals. The linear chromosome is again digested at specific site with two arms with different selection marker. Genomic insert is then ligated into YAC vector using DNA ligase enzyme. The recombinant vectors are transformed into yeast cells and screened for the selection markers to obtain recombinant colonies.
- Yeast expression vectors, such as YACs, YIPs (yeast integrating plasmids), and YEPs (yeast episomal plasmids), have advantageous over bacterial artificial chromosomes (BACs). They can be used to express eukaryotic proteins that require post-translational modification. However, YACs have been found to be less stable than BACs.

Some recombinant plasmids have the ability to incorporate multiple replication origins and other elements that allow them to be used in more than one species (for example, yeast or *E. coli*).

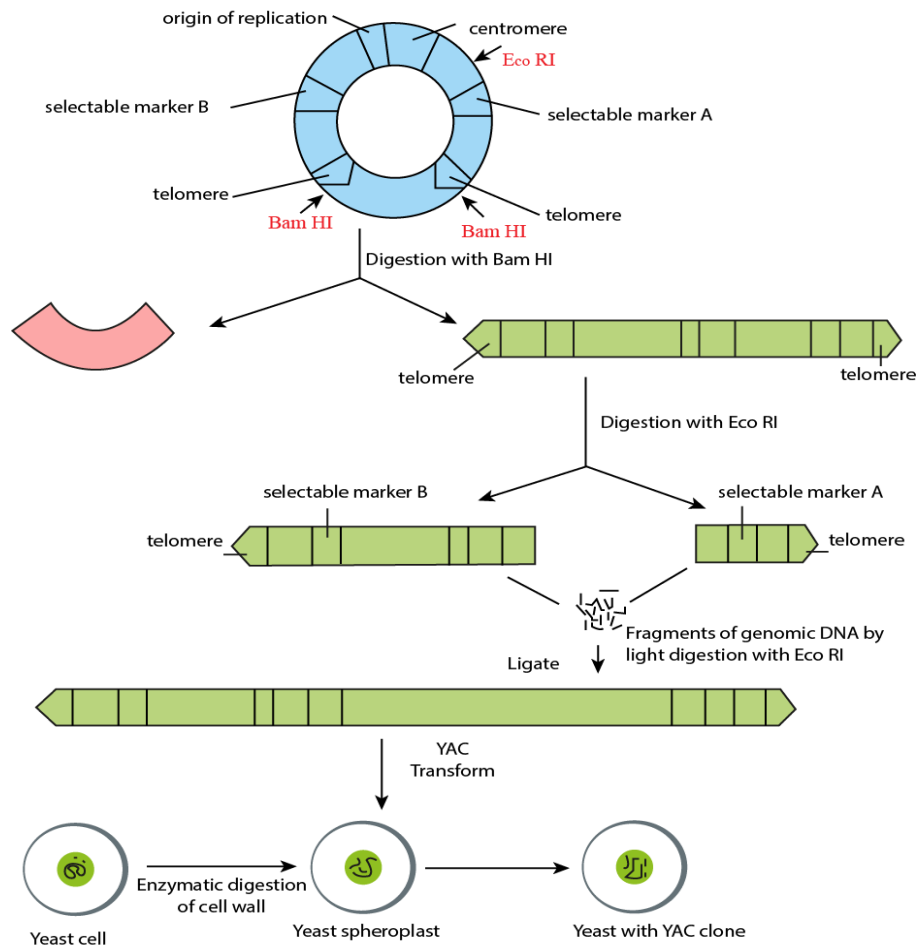


Fig 1-4.6.2: YAC vector

1-4.6.3 Mammalian Artificial Chromosomes: MAC

MACs or mammalian artificial chromosomes, like YACs, rely on the presence of centromeric and telomeric sequences and origin of DNA replication. They involve autonomous replication and segregation in mammalian cells, as opposed to random integration into chromosomes (as for other vectors). They can be modified for their use as expression systems of large genes, including not only the coding region but can contain control elements. Two principal procedures exist for the generation of MACs.

- 1) In one method, telomere-directed fragmentation of natural chromosomes is used. For example, a human artificial chromosome (HAC) has been derived from chromosome 21 using this method.
- 2) Another method involves *de novo* assembly of cloned centromeric, telomeric, and replication origins *in vitro*.

MAC vectors are difficult to assemble as compared to YAC vectors. Mammalian DNA has higher degree of repetition and larger centromere and telomere regions. Also the sequences necessary for chromosome replication in mammalian system are not well defined till now. MAC vectors have application in the field of gene therapy and eukaryotic protein expression and production.

Bibliography:

- Cooke H. 2001. Mammalian artificial chromosomes as vectors: progress and prospects; *Cloning Stem Cells*, 3(4): 243-249.
- Dente L., Cesareni G., Cortese R. 1983. pEMBL: a new family of single stranded plasmids; *Nuc. Acid Res.* 11 (6) 1645-1655.
- Hall B.G. 2004. Predicting the evolution of antibiotic resistance genes. *Nat Rev Microb* 2 (5): 430–435.
- <http://bioinfo2010.wordpress.com/2009/07/08/vector-bacteriophage-lambda-and-m13-7th-april/>
- Kim *et al.* 1992. Stable propagation of cosmid-sized human DNA inserts in an F-factor based vector. *Nucleic Acids Res.* **20** (5): 1083–1085.

- Kim *et al.* 1995. Construction and utility of a human chromosome 22-specific Fosmid library; *Genetic Analysis: Biomol Eng* **12** (2): 81–84.
- Qi H., Lu H., Qiu H.J., Petrenko V. et al. 2012. Phagemid Vectors for Phage Display: Properties, Characteristics and Construction; *J.Mol. Bio* (417), 129-143.
- Shizuya H. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA*, 89(18): 8794–8797.
- Tursun B, Cochella L, Carrera I, Hobert O. 2009. A Toolkit and Robust Pipeline for the Generation of Fosmid-Based Reporter Genes in *C. elegans*. *PLoS ONE* 4(3): e4625.

MODULE 1- LECTURE 5

SAFETY GUIDELINES FOR RECOMBINANT DNA RESEARCH

1-5.1 Introduction:

During early era of life sciences, biosafety principles and guidelines were mostly applied in the field of microbiology and medical practices. In recent time, biosafety guidelines are designed and applied to research involving recombinant DNA (rDNA) techniques. Handling, production, storing and transportation of genetically modified organism (GMOs) involve different biosafety issues under different category. Biosafety practices deal with the application of standard safety principles handling hazardous material/agents to minimize potential harmful effect on human health and environment.

The definition of biosafety corresponds to recombinant DNA technology can be described as: “Application of safety principles to laboratory practices in which potentially hazardous materials or organism are manipulated or handled.”

National Institutes of Health (NIH) developed certain guidelines on rDNA research in May, 1976. Similarly, Department of Biotechnology (Govt. of India) has also come out with its set of guidelines which is available at <http://dbtindia.nic.in/index.asp>. Scope of the guidelines suggested by DBT encompasses research, large scale operations and environmental risks. The guidelines prescribe specific actions that include establishing safety procedures for rDNA research, production and release to the environment and setting up containment conditions for certain experiments.

Biosafety regulatory principles and protocols regulates the potential risk and allow access to the benefits of rDNA technology. Risk assessment and risk management are two important component of biosafety.

1-5.2 Biosafety Levels (BSL):

All the facilities handling microorganisms and materials containing recombinant DNA molecules have risk assessment program. Depending on the risk possessed by the samples, four biosafety levels have been assigned to rDNA research facilities. Each BSL facility has requirement of unique design features and safety equipments.

1-5.2.1 Biosafety level-I (BSL-I):

- Agents: Characterized strains of microorganisms known to cause no disease in healthy adults. eg. *E. coli*, *S. cerevisiae*, *B. subtilis* etc.
- Recombinant DNA based research activities involving non-pathogenic micro-organisms for expression of genes using plasmid vectors or low risk viral vectors.
- Work practice: Standard aseptic microbiological techniques.
- Safety equipment requirement: Lab coats and eye protection recommended.
- Facilities: Bench top, sink etc.

1-5.2.2 Biosafety level-II (BSL-II):

- Agents: Handling of micro-organisms which possess moderate hazard to personal and environment.
- rDNA based research activities in micro-organisms using non-viral or viral vectors.
- Work practice: Standard BSL-I practices with addition of limited access, biohazard sign, defined procedure for disposal of “Regulated Medical Waste”, proper training to lab personal and medical surveillance.
- Safety equipment: Class-II biological safety cabinet, lab coats, gloves, eye/face protection, physical containment equipment to reduce infectious aerosol exposure or splashes.
- Facility: BSL-I facility with addition of autoclave, decontamination facility and proper airflow.

1-5.2.3 Biosafety level-III (BSL-III):

- Agents: Handling of micro-organisms which are designated as hazardous or potentially lethal agents to personal and environment.
- Laboratory personnel must have specific training in handling infectious micro-organisms and should be supervised by scientist competent in handling infectious agents.
- Work practices: BSL-2 practices, with the addition of: controlled access, on-site decontamination of all waste and lab clothing and medical surveillance.
- Safety equipment: Class-III biological safety cabinet, lab coats, gloves, eye/face protection, respiratory protection, physical containment equipment to reduce infectious aerosol exposure or splashes.
- Facility: BSL-III facility has specific criteria to meet. Lab should have double door entry with physical separation of working area from the access corridors, directional airflow in lab, and no recirculation of exhaust air in the lab, sufficient decontamination facility, in lab autoclave etc.

1-5.2.4 Biosafety level-IV (BSL-IV):

- Agents: Hazardous and potentially lethal organisms that possess high individual risk of laboratory transmitted disease for which there is no vaccine or treatment, or a related agent with unknown risk of transmission.
- Laboratory personnel must have specialized training in handling BSL-IV agents and should be supervised by scientist competent in handling infectious agents.
- Safety equipment: Class-IV biological safety cabinet, lab coats, gloves, eye/face protection, respiratory protection, physical and containment equipment to reduce infectious aerosol exposure or splashes.
- Facility: BSL-IV facility requires specialized design to minimize the exposure to risk and only the authorized entry should be permitted in laboratory area in BSL-IV labs.

1-5.3 Risk Analysis:

The foundation of any safety program is the use of control measures appropriate for the risk posed by the activities and the agents in use. The process of analyzing and determining the risk associated with recombinant DNA work is called as **Risk analysis**. The principle behind biosafety regulations is to minimize the risk to human health and safety, and the conservation of environment including safe handling of hazardous material. Risk analysis consists of three components: risk assessment, risk management and risk communication.

Risk Assessment: Estimation and determination of risk associated with the handling and production of a recombinant DNA molecule.

Risk Management: The process of analyzing possible prevention measures to minimize the risk and designing policies accordingly including implementation of them.

Risk Communication: The exchange of information and opinions on risk management between academic parties, industry, consumers and policy makers.

1-5.3.1 Risk Assessment:

The biosafety level is determined based on the risk associated with the work. The principle investigator is responsible for implementing the necessary safety requirements in his/her laboratory. Risk assessment process accounts the following criteria to determine biosafety level:

- i. **Pathogenicity** – The ability of an organism to cause disease in human system.
- ii. **Virulence** – The severity of the disease (lethal/non lethal, availability of cure etc) in a healthy adult.
- iii. **Proliferation** - the subsequent multiplication, genetic reconstruction, growth, transport, modification and die-off of these micro-organisms in the environment, including possible transfer of genetic material to other micro-organisms.
- iv. **Transmission route** - The possible route of transmission (mucous membrane, inhalation etc) to establish the disease in human or other organism.
- v. **Infectious dose (ID)** – The amount of infectious agent required to cause disease in healthy human.

- vi. **Antibiotic/disinfectant resistance** – The resistance acquired by the infectious agent to available antibiotic/disinfectant.

The risk associated with recombinant DNA technology can be categorized under different headings based on their implication on different platforms.

General Scientific Considerations-

- A. Characteristics of Donor and Recipient Organisms
- Taxonomy, identification, source and culture
 - Genetic characteristics of donor and recipient organisms
 - Pathogenic and physiological traits of donor and recipient organisms
- B. Properties of the modified/engineered organism
- C. Description of (a) modification, (b) nature, function and source of the insert, (c) vector construction, (d) transfer into host, (e) stability of insert, (f) frequency of mobilization, (g) rate and level of expression, and (h) Influence of the recipient organism on the activity of the foreign protein.

Human Health Considerations:

- A. Characteristics of the modified/engineered organism.
- Comparison of the recombinant organism to the wildtype organism regarding pathogenicity.
 - Transmission route to human.
 - Pathogenicity to humans (or to animals if appropriate).
- B. Health considerations generally associated with the presence of non-viable organisms or with the products of rDNA processes.
- C. Management of personnel exposure, including biological measures and physical and organizational measures.

Environmental and Agricultural Considerations:

- A. Ecological traits relating to the donor and recipient environment.
- B. Properties of environment where the engineered organism are being applied.
- C. Survival, multiplication and dissemination of the engineered organism in the environment.

- D. Interactions of engineered organism(s) with biological systems (target and non-target populations, stability, and routes of dissemination).
- E. Potential environmental impacts (Effect on target and non-target organisms and ecosystems).

1-5.3.2 Risk Management:

Risk management in biosafety issues is related to the target site where the practice is conducting (laboratory, industry, agriculture field etc.).

Recommendations: **General**

- i. Harmonization of approaches to rDNA techniques can be facilitated by exchanging principles or guidelines for national regulations; developments in risk analysis; and practical experience in risk management. Therefore, information should be shared as freely as possible.
- ii. There is no scientific basis for specific legislation for the implementation of rDNA techniques and applications. Member countries should examine their existing oversight and review mechanisms to ensure that adequate review and control may be applied while avoiding any undue burdens that may hamper technological developments in this field.
- iii. Any approach to implement guidelines should not impede future developments in rDNA techniques. International harmonization should recognize this need.
- iv. To facilitate data exchange and minimize trade barriers between countries, further developments such as testing methods, equipment design, and knowledge of microbial taxonomy should be considered at both national and international levels. Due account should be taken of ongoing work on standards within international organizations.
- v. Special efforts should be made to improve public understanding of the various aspects of rDNA techniques.
- vi. For rDNA applications in industry, agriculture and the environment, it will be important for member countries to watch the development of these techniques. For certain industrial applications and for environmental and agricultural

applications of rDNA organisms, some countries may wish to have a notification scheme.

- vii. Recognizing the need for innovation, it is important to consider appropriate means to protect intellectual property and confidentiality interests while assuring safety.

Recommendations: Specific for Industry

- i. The large-scale industrial application of rDNA techniques wherever possible should utilize microorganisms that are intrinsically of low risk. Such microorganisms can be handled under conditions of Good Industrial Large-Scale Practice (GILSP).
- ii. If a recombinant microorganism cannot be handled merely by GILSP, measures of containment corresponding to the risk assessment should be used in addition to GILSP.
- iii. Further, research to improve techniques for monitoring and controlling non-intentional release of rDNA organisms should be encouraged in large-scale industrial applications requiring physical containment.

1.5.4 Containment levels:

Biosafety containment levels have to be designated for a facility depending on the level of risk associated with the biological and chemical agents used and released from it. Following NIH (National Institute of Health, USA) and DBT (Department of Biotechnology, India) guidelines, different facilities for biological research have been classified under three containment levels.

1.5.4.1 Containment Category 1:

- Viable organisms should be handled in a production system which physically separates the process from the environment;
- Exhaust gases should be treated to minimize (i.e. to reduce to the lowest practicable level consistent with safety) the release of viable organisms;
- Sample collection, addition of materials to the system and the transfer of viable organisms to another system should be done in a manner which minimizes release;
- Bulk quantities of culture fluids should not be removed from the system unless the viable organisms have been inactivated by validated means;
- Effluent from the production facility should be inactivated by validated means prior to discharge.

1.5.4.2 Containment Category 2:

- Viable organisms should be handled in a production system which physically separates the process from the environment;
- Exhaust gases should be treated to prevent the release of viable organisms;
- Sample collection, addition of materials to a closed system and the transfer of viable organisms to another closed system should be done in a manner which prevents release;
- Culture fluids should not be removed from the closed system unless the viable organisms have been inactivated by validated chemical or physical means;
- Seals should be designed to prevent leakage or should be fully enclosed in ventilated housings;
- Closed systems should be located in an area controlled according to the requirements;
- Effluent from the production facility should be inactivated by validated chemical or physical means prior to discharge.

1.5.4.3 Containment Category 3:

- Viable organisms should be handled in a production system which physically separates the process from the environment;
- Exhaust gases should be treated to prevent the release of viable organisms;
- Sample collection, addition of materials to a closed system and the transfer of viable organisms to another closed system should be done in a manner which prevents release;
- Culture fluids should not be removed from the closed system unless the viable organisms have been inactivated by validated chemical or physical means;
- Seals should be designed to prevent leakage or should be fully enclosed in ventilated housings;
- Production systems should be located within a purpose built controlled area according to the requirements;
- Entry should be restricted in the laboratory area and only persons with appropriate authority should be allowed access to the working area.

Effluent from the production facility should be inactivated by validated chemical or physical means prior to discharge.

Different containment levels have been assigned for rDNA GILSP (Good industrial large scale practice) micro-organisms. Examples of containment approaches for recombinant organisms are discussed below:

S.No.	Specifications	Containment categories		
		1	2	3
1.	Viable organisms should be handled in a system which physically separates the process from the environment (closed system)	Yes	Yes	Yes
2	Exhaust gases from the closed system should be treated so as to:	Minimum release	Prevent release	Prevent release
3.	Sample collection, addition of materials to a closed system and transfer of viable organisms to another closed system, should be performed so as to:	Minimum release	Prevent release	Prevent release
4.	Bulk culture fluids should not be removed from the closed system unless the viable organisms have been:	Inactivated by validated means	Inactivated by validated chemical or physical means	Inactivated by validated chemical or physical means
5.	Seals should be designed so as to:	Minimum release	Prevent release	Prevent release
6.	Closed systems should be located within a controlled area	Optional	Optional	Yes, and purpose-built
a) Biohazard signs should be posted b) Access should be restricted to nominated personnel only c) Personnel should wear protective clothing		Optional	Yes	Yes
		Optional	Yes	Yes, via an air-lock
		Yes, work clothing	Yes	Yes, specially designed for such

				facilities to provide protection
	d) Decontamination and washing facility should be provided for personnel	Yes	Yes	Yes
	e) Personnel should shower before leaving the controlled area	No	Optional	Yes
	f) Effluent from sinks and showers should be collected and inactivated before release	No	Optional	Yes
	g) The controlled area should be adequately ventilated to minimize air contamination	Optional	Optional	Yes
	h) The controlled area should be maintained at an air pressure negative to atmosphere	No	Optional	Yes
	i) Input air and extract air to the controlled area should be HEPA filtered	No	Optional	Yes
	j) The controlled area should be designed to contain spillage of the entire contents of the closed system	No	Optional	Yes
	k) The controlled area should be sealable to permit fumigation	No	Optional	Yes
7.	Effluent treatment before final discharge	Inactivated by validated means	Inactivate d by validated or physical means	Inactivate d by validated or physical means

Table 1.5.4: Containment levels for different facilities.

Recommendations: Specific for Environment and Agriculture

- Considerable data on the effects of different micro-organisms on environmental and human health exist in literature and should be used to guide risk assessments.
- It is important to evaluate rDNA organisms for potential risk prior to applications in agriculture and the environment. However, the development of general international guidelines governing such applications is premature at this time. An independent review of potential risks should be conducted on a case-by-case basis prior to the application.
- Development of organisms for agricultural or environmental applications should be conducted in a stepwise fashion, moving, where appropriate, from the laboratory to the growth chamber and greenhouse, to limited field testing and finally, to large-scale field testing.
- Further research to improve the prediction, evaluation and monitoring of the outcome of applications of rDNA organisms should be encouraged.

Further Reading

- RECOMBINANT DNA SAFETY CONSIDERATIONS. Published by Department of Biotechnology, Ministry of Science and Technology, Govt. of India. Available at Department of Biotechnology. *website: <http://dbtindia.nic.in/index.asp>*
- RECOMBINANT DNA SAFETY GUIDELINES 1990. Published by Department of Biotechnology, Ministry of Science and Technology, Govt. of India. Available at Ministry of Environment & Forests. *website: <http://moef.nic.in/index.php>*

MODULE 1- LECTURE 6

CONTROL OF SPILLS AND MECHANISM OF IMPLEMENTATION OF BIOSAFETY GUIDELINES

1-6.1 Introduction:

National Institute of Health (NIH) has developed certain management procedures that can be adopted by research workers themselves to contain minor spills involving small quantities of biological materials/ experimental samples related to recombinant DNA. However, management of decontamination and cleanup of large spill or a spill involving a highly infectious agent are best left to the biosafety officer of the laboratory. It is the responsibility of the principal investigator to maintain an enough supply of a chemical disinfectant effective against the investigated microorganisms.

A spill may happen under various circumstances as below-

- in a biological safety cabinet,
- in the open laboratory,
- in a centrifuge,
- biological spill on a person and
- spill involving radioactive materials.

In each case, before starting the procedure and during the procedure, protective gloves, lab coat or gown, and eye protection should be worn. Researchers in the vicinity and the principal investigator should also be warned immediately. Decontamination with a recommended disinfectant procedure should be followed. Similarly, rDNA and transgenic organisms must be treated the same as medical or infectious waste before disposal. All wastes resulting from the procedure are to be disposed in the designated biowaste container.

1-6.2 Spill in a Biological Safety Cabinet:

A spill that is confined within a biological safety cabinet generally presents little or no hazard to personnel in the area. Chemical disinfection procedures are to be initiated at once while the cabinet continues to operate. The disinfectant should be active against the organisms of potential hazard.

- i. Spray or wipe the walls, work surfaces, and equipment with the chosen disinfectant. Disinfectants with detergent have the advantage of detergent activity that will help clean the surfaces by removing both dirt and microorganisms.
- ii. Minimize the generation of aerosols and use sufficient disinfectant to ensure that drain pans and catch basins below the work surface contain disinfectant. The front exhaust shall also be wiped and the disinfectant drained into a container.
- iii. Lift the front exhaust grill and tray and wipe all surfaces. Wipe the catch basin and drain the disinfectant into a container.

This procedure will not disinfect the filters, fans, air ducts, and other interior parts of the cabinet. Contact bio safety officer if the interior cabinet needs to be disinfected.

1-6.3 Spill in the Open Laboratory:

For a spill in the open laboratory outside a biological safety cabinet, the spill response depends on the size of the spill and hazard of the material.

- i. A minimally hazardous material spilled without generating appreciable aerosols can be cleaned with a paper towel soaked in a chemical disinfectant.
- ii. A spill of a larger volume of hazardous material with aerosol generation requires evacuation of the room, waiting for aerosol reduction, donning personal protective gear (including appropriate respiratory protection), selecting a disinfectant effective against the organisms involved, and cleaning as described above.

Following cleanup, responsible personnel shall wash or shower the infected area with a disinfectant soap.

- iii. If clothing gets contaminated, it should be carefully removed and folded the contaminated area inward and keep it for autoclaving.
- iv. Wash arms, face, and hands with disinfectant.

1-6.4 Spill in a Centrifuge:

A biological spill in a centrifuge has the potential for producing large volumes of aerosols.

- i. On becoming aware that a spill may have occurred within a centrifuge or other piece of equipment, turn off the equipment, allow aerosols to settle, and then decontaminate it.
- ii. Place contaminated equipment in a leak proof bag and move it to a biological safety cabinet, if possible for decontamination.

1-6.5 Biological Spill on a Person:

If a biological material is spilled on a person, emergency response is based on the hazardous level of the biological agent spilled, the amount of material spilled, and whether significant aerosols were generated.

- i. If aerosol formation is believed to have been associated with the spill, a contaminated person shall leave the contaminated area immediately. If possible, he/she should go to another laboratory area so that hallways and other public areas do not become contaminated.
- ii. Contaminated clothes should be removed and placed in red or orange biohazard bags for disinfecting. Contaminated skin shall be flushed with water followed by washing with a disinfectant soap. Showering can also be done, depending on the extent of the spill. A designated shower room should be earmarked near the laboratory.

1-6.6 Spill Involving Radioactive Materials:

Spills involving radioactive materials should be strictly handled by safety officer in charge. Following safety precautions should be taken under such situation:

- i. Other laboratory personals should be warned immediately and if possible the place should be vacated.
- ii. The principal investigator should be informed about the spill.
- iii. The institute biosafety officer should be contacted immediately.
- iv. One should not try to handle anything by himself.

1-6.7 Mechanism of Implementation of Biosafety Guidelines:

The guidelines suggest compliance of the safeguards through voluntary as well as regulatory approach. The implementation is enforced through the development of institutional frame work of advisory and regulatory bodies to deal with the specific and discretionary actions on the following:

- i. Self regulation and control in the form of guidelines on recombinant research activities; and
- ii. Regulation of large scale use of engineered organisms in production activity and release of organisms in environmental applications under statutory provisions.

The institutional mechanism as proposed for implementation of guidelines is shown in Figure 1-6.7.

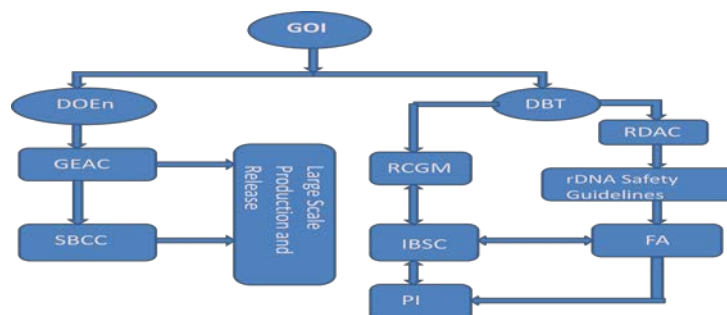


Fig 1-6.7. Implementation of Biosafety guidelines

(**GOI**: Govt. of India, **DBT**: Department of Biotechnology; **RDAC**: Recombinant DNA Advisory Committee; **IBSC**: Institutional Biosafety Committee; **RCGM**: Review Committee on Genetic Manipulation; **DOEn**: Department of Environment; **GEAC**: Genetic Engineering Approval Committee; **SBCC**: State Biotechnology Coordination Committee; **PI**: Principle Investigator (R&D/Industry/Others); **FA**: Funding Agency (Govt./Private or Public Institution))

Mainly it consists of the following:-

- i. Recombinant DNA Advisory Committee (RDAC)
- ii. Institutional Biosafety Committee (IBSC)
- iii. Review Committee on Genetic Manipulation (RCGM)
- iv. Genetic Engineering Approval Committee (GEAC)

1-6.7.1 Recombinant DNA Advisory Committee (RDAC)

The Committee should take note of development and advances at national and international levels in Biotechnology towards the correctness of the safety regulation for India on recombinant research use and applications. The committee should be aware of recent developments and advances in safety regulation in recombinant research at both national and international levels. It should meet once in 6 months or sooner for this purpose.

The specific terms of reference for Recombinant Advisory Committee include the following:

- i) To evolve long term policy for research and development in Recombinant DNA research.
- ii) To formulate the safety guidelines for Recombinant DNA Research to be followed in India.

iii) To recommend type of training program for technicians and research fellows for making them adequately aware of hazards and risks involved in recombinant DNA research and methods for avoiding it.

1-6.7.2 Institutional Biosafety Committee (IBSC)

Institutional Biosafety Committee (IBSC) is to be constituted in all centres engaged in genetic engineering research and production activities. The Institutional Biosafety Committee shall be the nodal point for interaction within institution for implementation of the guidelines. Any research project which is likely to have biohazard potential (as envisaged by the guidelines) during the execution stage or which involve the production of either microorganisms or biologically active molecules that might cause bio-hazard should be notified to IBSC.

The biosafety functions and activity include the following:

- i. Registration of Bio-safety Committee membership composition with RCGM and submission of reports: IBSC will provide half yearly report on the ongoing projects to RCGM regarding the observance of the safety guidelines on accidents, risks and on deviations if any. A computerized Central Registry for collation of periodic report on approved projects will be set up with RCGM to monitor compliance on safeguards as stipulated in the guidelines.
- ii. Review and clearance of project proposals falling under restricted category that meets the requirements under the guidelines: IBSC should make efforts to issue clearance quickly on receiving the research proposals from investigators.
- iii. Tailoring biosafety program to the level of risk assessment.
- iv. Training of personnel on biosafety.
- v. Instituting health monitoring program for laboratory personnel.
- vi. Adopting emergency plans.

1-6.7.3 Review Committee on Genetic Manipulation (RCGM)

The RCGM will have the following functions:

- i. To establish procedural guidance manual - for regulatory process involving genetically engineered organisms in research, production and applications related to environmental safety.
- ii. To review the reports in all approved ongoing research projects involving high risk category and controlled field experiments and ensures that safeguards are maintained as per guidelines.
- iii. To recommend the type of containment facility and the special containment conditions to be followed for experimental trials and for certain experiments.
- iv. To advise customs authorities on import of biologically active material, genetically engineered substances or products and on excisable items to Central Revenue and Excise.
- v. To assist Department of Industrial Development, Banks towards the clearance of applications in setting up industries based on genetically engineered organisms.
- vi. To assist the Bureau of Indian Standards to evolve standards for biologics produced by recombinant DNA (rDNA) technology.
- vii. To advise on intellectual property rights with respect to rDNA technology on patents.

1-6.7.4 Genetic Engineering Approval Committee (GEAC)

Genetic Engineering Approval Committee (GEAC) will function under the Department of Environment (DOEn) as statutory body for review and approval of activities involving large scale use of genetically engineered organisms and their products in research and development, industrial production, environmental release and field applications.

The functions include giving approval from environmental view on:

- i. Import, export, transport, manufacture, process, selling of any microorganisms or genetically engineered substances or cells including food stuffs and additives that contains products derived by Gene Therapy.

- ii. Discharge of Genetically engineered/classified organisms/cells from Laboratory, hospitals and related areas into environment.
- iii. Large scale use of genetically engineered organisms/classified microorganisms in industrial production and applications. (Production shall not be commenced without approval).
- iv. Deliberate release of genetically engineered organisms. The approval will be for a period of 4 years.

The funding agency will be responsible for approval and clearing of research proposals for grants in aid in respect of rDNA research activities. The funding agency at the centre and state level will be advised to ensure that the guidelines are taken into account for compliance while supporting grants on research projects. Investigators will be required to submit as part of the project application an evaluation of biohazards that may arise and also the requirement on the type of containment facility, certified by IBSC. The funding agency should state clearly that support on approved projects will be withdrawn in case of deliberate violation or avoidable negligence of the rDNA guidelines. The investigators will also be asked to make a declaration in their publications that the work was carried out following the national guidelines. The funding agency will annually submit to RCGM the list of approved projects that come under high risk categories.

Website

- Department of Biotechnology, India; *Guidelines and handbook for Institutional Biosafety Committees (IBSCs)* (2nd revised edition); 2011. Website: <http://dbtbiosafety.nic.in/>